ECONOMETRIC INFERENCE IN MODELS WITH NONSTATIONARY TIME SERIES

Michalis P. Stamatogiannis

Thesis submitted to the University of Nottingham
for the degree of Doctor of Philosophy

July 2010

ABSTRACT

Econometric Inference in Models With Nonstationary Time Series

Michalis P. Stamatogiannis

We investigate the finite sample behaviour of the ordinary least squares (OLS) estimator in vector autoregressive (VAR) models. The data generating process is assumed to be a purely nonstationary first-order VAR. Using Monte Carlo simulation and numerical optimization we derive response surfaces for OLS bias and variance in terms of VAR dimensions both under correct model specification and under several types of over-parameterization: we include a constant, a constant and trend, and introduce excess autoregressive lags. Correction factors are introduced that minimise the mean squared error (MSE) of the OLS estimator. Our analysis improves and extends one of the main finite-sample multivariate analytical bias results of Abadir, Hadri and Tzavalis (1999), generalises the univariate variance and MSE results of Abadir (1995) to a multivariate setting, and complements various asymptotic studies.

The distribution of unit root test statistics generally contains nuisance parameters that correspond to the correlation structure of the innovation errors. The presence of such nuisance parameters can lead to serious size distortions. To address this issue, we adopt an approach based on the characterization of the class of asymptotically similar critical regions for the unit root hypothesis and the application of two new optimality criteria for the choice of a test within this class. The correlation structure of the innovation sequence takes the form of a moving average process, the order of which is determined by an appropriate information criterion. Limit distribution theory for the resulting test statistics is developed and simulation evidence suggests that our statistics have substantially reduced size while retaining good power properties.

Stock return predictability is a fundamental issue in asset pricing. The conclusions of empirical analyses on the existence of stock return predictability vary according to the time series properties of the economic variables considered as potential predictors. Given the uncertainty about the degree of persistence of these variables, it is important to operate in the most general possible modelling framework. This possibility is provided by the IVX methodology developed by Phillips and Magdalinos (2009) in the context of cointegrated systems with no deterministic components. This method is modified in order to apply to multivariate systems of predictive regressions with an intercept in the model. The resulting modified IVX approach yields chi-squared inference for general linear restrictions on the regression coefficients that is robust to the degree of persistence of the predictor variables. In addition to extending the class of generating mechanisms for predictive regression, the approach extends the range of testable hypotheses, assessing the combined effects of different explanatory variables to stock returns rather than the individual effect of each explanatory variable.

Acknowledgements

I am grateful to my supervisors, Professor Rob Taylor and Dr. Tassos Magdalinos. Professor Taylor's invaluable insight and expediency, together with Dr. Magdalinos' technical guidance and unparalleled support, helped me to substantially improve the quality of my research. Not only have I learnt a great deal from them, but this thesis would not have been possible without their assistance.

I would like to thank Professor Karim Abadir who made extensive comments on parts of this thesis.

Academic collaboration with Dr. Steve Lawford provided me with valuable experience and motivation for research. I also extend a special thank you to him for reading my drafts and supporting me every step of the way.

Working with Dr. Alexandros Kostakis aroused my interest in financial econometrics. His in-depth knowledge on the subject has been very motivating.

Professor Steve Leybourne and Professor Jean-Yves Pitarakis made insightful comments, acting as examiners for this thesis.

I would also like to thank Dr. David Harvey who was always willing to help me.

I am grateful to the Department of Economics and Econometrics at the University of Groningen, and personally its Head Professor Ruud Koning, for offering me a position at such an early stage of my career.

Further, I must extend gratitude to my father Panagiotis for his support in my studies. He instilled in me a passion for learning and professional betterment.

Finally, my warmest and deepest thanks go to my dearest mother, Eleni, for her love, patience and encouragement in every area of my life.

Author's Declaration

Chapter 2 of this thesis is based on a paper written by Steve Lawford and myself, which has been published in the Journal of Econometrics (Lawford and Stamatogiannis, 2009). Chapter 4 of this thesis is based on a collaboration between Alexandros Kostakis, Tassos Magdalinos and myself. Chapter 3 has resulted from research conducted by the present author exclusively.

Contents

CHAPTER 1

Introduction

Nonstationarity has dominated the time series literature for the last three decades. Interest in the topic was initiated by the empirical relevance of nonstationary data. Fundamental issues in macroeconomics and finance such as growth, the efficient market hypothesis and business cycles are crucially influenced by the existence of nonstationarity. Nelson and Plosser (1982) found that many macroeconomic series had a unit root and discussed the implications of such form of nonstationarity on the theory of business cycles. The impact of nonstationary variables to both economic theory and economic forecasting created the need for statistical methods that would detect the persistence properties of economic time series and provide valid inference in cases where these series exhibit stochastic trends.

The inadequacy of standard inference when applied to nonstationary series was exposed by Granger and Newbold (1974) who introduced the idea of spurious regression. Using Monte Carlo simulations, they showed that standard regression methods can provide statistical evidence in favour of fallacious relationships among variables that contain unit roots. Granger (1981) formalised the notion of a meaningful relationship among nonstationary data series by introducing the definition of *co-integration*. Phillips (1986) developed asymptotic methods that explained analytically the problems associated with spurious regressions and provided the correct regression theory. His elegant method, based on a functional central limit theorem and the preservation of weak convergence by continuous mappings, provided the foundation for a huge research programme on formal econometric inference for nonstationary processes that attracted a large number of econometricians, statisticians and probabilists. The topics discussed in this thesis relate to the econometric estimation and testing in the presence of various forms of nonstationarity.

Chapter 2 investigates certain finite sample properties of ordinary least squares (OLS) estimation in vector autoregressive (VAR) models. Assuming a data generating process of the form

$$\begin{aligned} x_t &= Rx_{t-1} + \varepsilon_t; \quad t = 1, 2, \ldots T, \\ R &= I_k, \end{aligned} \tag{1.1}$$

where I_k the $k \times k$ identity matrix and $\{\varepsilon_t\}_1^T$ is a sequence of independent and identically distributed normal random vectors with mean 0 and positive definite covariance matrix Ω, the OLS estimator is given by

$$\widehat{R} = \sum_{t=1}^{T} x_t x_{t-1} \left(\sum_{t=1}^{T} x_{t-1} x'_{t-1} \right)^{-1}. \tag{1.2}$$

The main contribution of Chapter 2 is the characterisation of the finite sample behaviour of OLS estimators in relation to the VAR dimension and autoregressive lag length misspecification in (1.1). We provide expressions for the OLS bias and variance and derive correction factors that deliver minimum mean squared error (MSE) estimators.

The finite sample properties of the OLS estimator $\widehat{R}$ in (1.2) in the univariate case ($k = 1$) has been the subject of considerable study. MacKinnon and Smith (1998) show that $\widehat{R}$ is negatively biased for $R = 1$, with bias decreasing with the sample size and positively biased for certain values of R (notably $R = -1$). They also show that the bias function for $\widehat{R}$ is almost linear for $R \in [-0.85, 0.85]$ and highly nonlinear for values of R close to 1 and -1. The bias and exact moments of the OLS estimator in autoregressive models have been discussed recently by *inter alia* Nankervis and Savin (1988), Tsui and Ali (1989, 1994), Vinod and Shenton (1996) and Gonzalo and Pitarakis (1998). Abadir (1993) derives a high-order closed form approximation of the finite sample bias of $\widehat{R}$ with $|R| = 1$:

$$b \equiv \mathrm{E}\left(\widehat{R}\right) - R = \frac{\sqrt{2}}{T}\mu_T, \tag{1.3}$$

where $\mu_T = \mathrm{E}\left[T\left(\widehat{R} - R\right)/\sqrt{2}\right]$. Abadir (1993) shows that exponential functions in polynomials of T^{-1} may be used to describe the bias. A heuristic process (5 datapoints, no diagnostics reported) gives the simple approximation $\mu_T \approx \mu_\infty \exp\left(-2.6138\, T^{-1}\right)$, where μ_T and μ_∞ are exact values from Evans and Savin (1981, p. 769, Table III). The univariate bias approximation is obtained from (1.3) and an OLS regression of $\ln(\mu_T/\mu_\infty)$ on $1/T$ as

$$b \approx -1.7814\left(\frac{1}{T}\right)\exp\left(\frac{-2.6138}{T}\right), \tag{1.4}$$

where -1.7814 is the expected value of the limiting distribution of $T\left(\widehat{R}-1\right)$.

OLS bias in VAR models with $k \geq 1$ has been studied by Abadir, Hadri and Tzavalis (1999) (hereafter referred to as AHT). This Chapter shows that the bias of $\widehat{R}$ in (1.2) is a scalar matrix, (i.e. a diagonal matrix with equal diagonal elements) and is not a function of Ω. In particular, the bias matrix is approximately equal to the dimension of the VAR times the univariate bias formula:

$$b^{AHT} \approx -1.7814\left(\frac{k}{T}\right)\exp\left(\frac{-2.6138}{T}\right)I_k, \tag{1.5}$$

for $T > k+2$, i.e. the bias is proportional to the dimension k of the VAR model, irrespective of the innovation covariance matrix Ω.

Chapter 2 extends the results of AHT, studying the finite sample properties of OLS bias for a data series generated by (1.1). We introduce over-parameterization in two directions: addition of deterministic components and addition of multiple autoregressive lags in the VAR model. Hence, while the process x_t is generated by (1.1), the estimated model is given by

$$x_t = \widetilde{\mu} + \widetilde{\delta}t + \widetilde{R}x_{t-1} + \sum_{j=1}^{p-1}\widetilde{\Gamma}_j\Delta x_{t-j} + \widetilde{\varepsilon}_t.$$

We extend the AHT scalar bias matrix result to OLS estimates obtained by the above overparametrized model, proving that the matrix $E(\widetilde{R}-R)$ is scalar. We find that the effect of the overparameterization causes the absolute value of finite sample bias to increase. An extensive simulation study yields estimated response surfaces for bias as a function of sample size, VAR dimension and VAR lag length. We also estimate response surfaces for the variance of the OLS estimator. Combining the information drawn from the response surfaces for OLS bias and OLS variance we compute correction factors that lead to minimum MSE estimators.

In Chapter 3 we derive test statistics for the unit root hypothesis that control size in the presence of autocorrelation in the error term and have comparatively good power properties. Early work of Fuller (1976), Dickey and Fuller (1979, 1981) and Said and Dickey (1984) led to "augmented" versions of unit root tests that take into account possible autocorrelation in the innovation errors of the model. Phillips and Perron (1988) proposed a nonparametric unit root test which allows for a very wide class of innovations, namely stationary (short memory) linear processes. Correlation is not assumed to have a specific parametric structure and is estimated by a nonparametric estimator of the spectral density function at zero frequency. All the above (and most subsequent) work treats error autocorrelation as a nuisance parameter that appears on the null asymptotic distribution of unit root test statistics and hence affects the size of unit root tests. This issue was highlighted by the numerical study of Schwert (1989) which demonstrated high size distortion of the ADF and Phillips Perron tests. DeJong *et al.* (1992) pointed out the problem of low power of unit root tests against trend-stationary alternatives. Dufour and King (1991) and Elliott *et al.* (1996) proposed local GLS detrending of the data in order to increase the power of the Dickey Fuller statistic. Ng and Perron (2001) use GLS detrended data to derive modified test statistics and modified information criteria for the determination of the truncation lag.

Chapter 3 presents an alternative approach based on the derivation of asymptotically similar unit root test statistics. Similarity refers to tests whose size is independent of nuisance parameters, in this case error autocorrelation. The characterisation of the class of similar tests in the context of autocorrelated errors is achieved using the methodology developed in Hillier (1987). Test statistics are selected from within the class of similar tests using two different optimality criteria: Bounded Norm Minimising (BNM) and Bounded Estimated Point Optimal (BEPO). These optimality criteria have been applied by Forchini and Marsh (2000) for the derivation of similar unit root tests under independence. In Chapter 3 we start from a uniformly most powerful critical region that accommodates correlated innovation errors that take the form of an MA(m) process. The BNM and BEPO optimality criteria are applied to choose statistics from the class of asymptotically similar tests. Due to the lack of a sufficient statistic for the estimation of the MA parameters, we estimate these parameters using maximum likelihood. The order of the MA component is determined by the use of information criteria. The asymptotic distributions of the resulting test statistics are derived for the case where the deterministic component of the model includes an intercept or an intercept and a linear trend. Subsequent numerical study shows that the BNM and BEPO statistics perform well relative to the other unit root tests in terms of both size distortion and power in finite samples. A feature that further distinguishes the BNM and BEPO statistics derived in Chapter 3 is that they do not suffer from the problem of *power reversal*. This term was introduced in the literature by Seo (2006) to describe a decrease in power as the true value of the parameter moves away from the null hypothesis[1].

Our simulation study reveals another problem that arises with test statistics that employ the modified information criteria (MIC) proposed by Ng and Perron (2001): for a given alternative value of the parameter of interest, there are cases when power decreases as the sample size increases. In other words, additional information leads to distorted inference which suggests that the tests are not consistent. The problem is related to a singular feature of MIC relative to traditional information criteria, namely the imposition of the null hypothesis. This offers excellent control over size. However, as the true value of the parameter of interest is moving away from the null, maintaining the null hypothesis through the MIC on the statistics can have a detrimental effect on the power of the associated tests.

The unit root tests derived in Chapter 3 do not suffer from the aforementioned problems. The power of the BNM and BEPO statistics increases as the true value of the autoregressive parameter moves farther away from the null hypothesis value. For a given alternative value of the parameter of interest, the power of these

[1]Surprisingly, the statistics resulting from Seo's procedure suffer from the same problem.

statistics increases as the sample size increases. Additionally, the BNM and BEPO tests appear to have relatively low size and high power compared to the statistics proposed by Ng and Perron (2001), Perron and Qu (2007) and Seo (2006).

In Chapter 4 we discuss inference in a broader framework of nonstationarity. As is often emphasised in applied work, economic and financial time series seem to exhibit persistence characteristics that do not always conform to the I(0)-I(1) dichotomy. In practice this means that applied researchers wish to model persistence in cointegrating regressions through series that have autoregressive roots in a general neighbourhood of unity. Considering persistent regressors that are not necessarily unit root processes is of particular importance for assessing the predictive power of economic and financial variables on stock returns. To this end, a well developed literature (Cavanagh *et al.*, 1995; Torous *et al.*, 2004; Campbell and Yogo, 2006) considers predictive regressions with local to unity regressors.

Accommodating such a generalisation, however, cannot be accomplished by standard methods. As Elliott (1998) showed, conventional cointegration methods such as fully modified OLS and dynamic OLS methods (Phillips and Hansen, 1990 and Stock and Watson, 1993 respectively) do not produce valid asymptotic inference in cases where the regressors have roots that are local to unity. Local to unity processes induce additional endogeneity that cannot be removed by standard methods. Similar problems occur when the regressors exhibit less persistence than local to unity processes. Such "mildly integrated" regressors were introduced by Phillips and Magdalinos (2007) and Giraitis and Phillips (2006). Given this wide class of possible generating mechanisms, there is a need to develop more robust approaches to estimation and inference that do not rely upon knowledge of the precise form of regressor persistence.

We apply the IVX method of Phillips and Magdalinos (2009) to the problem of testing for stock return predictability. The procedure is generalised by including an intercept in the model and provides robust inference in the following system of predictive regressions:

$$y_t = \mu + Ax_{t-1} + u_{0t}, \tag{1.6}$$

$$x_t = R_T x_{t-1} + u_{xt}, \tag{1.7}$$

$$R_T = I_k + C/T^{\alpha}, \quad \text{for some} \quad \alpha > 0 \tag{1.8}$$

for $t \in \{1, ..., T\}$, an $m \times k$ coefficient matrix A and innovations u_{0t}, u_{xt} that take the form of a stationary short memory linear process. The matrix C can be either zero or negative definite and together with α determine the degree of regressor persistence induced by the autoregressive matrix R_T in (1.8). If either $C = 0$ or $\alpha > 1$ in (1.8) the regressor x_t behaves as a unit root process. If $C < 0$ and $\alpha = 1$ the regressor in (1.7) is a local to unity process. If $C < 0$ and $\alpha \in (0, 1)$ the regressor belongs to the class of less persistent, mildly integrated processes introduced by Phillips and Magdalinos (2007).

Least squares limit theory for multivariate systems with mildly integrated regressors was established in Magdalinos and Phillips (2009). Phillips and Magdalinos (2009) employ a new instrumental variables procedure for the estimation of the coefficient matrix A in a cointegrated system. The idea is to construct instruments from the regressors by means of a suitable filtering. The approach is called "IVX estimation" because instruments are generated from the regressors by means of data differencing without using any external information. The degree of persistence of each IVX instrument is explicitly controlled so that the process is mildly integrated. This approach eliminates the local and moderate to unity endogeneity and produces a mixed normal limit distribution for the IVX estimator and standard chi-squared inference for restrictions on A irrespective of the degree of persistence of the regressors.

The contribution of Chapter 4 is twofold: First, motivated by the requirements of applied literature, the IVX methodology of Phillips and Magdalinos (2009) is extended to the case where an intercept is included in the model. The IVX estimator and the associated Wald test statistic are further modified to take into account the contemporaneous structure of predictive regressions. Second, an empirical analysis of the issue of predictability of stock returns is conducted by using the modified IVX methodology.

Apart from its robustness to the time series properties of the data generating process, the IVX methodology accommodates joint inference in the system (1.6)-(1.7), i.e. offers the possibility of assessing the predictive power of combinations of explanatory variables, or assessing the predictive power of a single regressor on multiple portfolios. This addresses a crucial empirical issue that could not be taken into account by previous studies on stock return predictability based on a local to unity framework (Cavanagh *et al.*, 1995; Torous *et al.*, 2004; Campbell and Yogo, 2006) because of the problems associated with multidimensional confidence interval construction for C. These problems do not affect IVX inference which is based on an endogeneity correction rather than Bonferroni type confidence intervals.

In the empirical part of Chapter 4 that assesses the predictability of the market portfolio, we use explanatory variables that are commonly employed as potential predictors. The market portfolio is decomposed to subcategories firstly according to the stocks' market value and secondly according to the stocks' book to market value. This categorisation of the market portfolio allows us to investigate whether the regressors predict specific subcategories of the market portfolio and also whether a regressor can jointly predict the subcategories of the market portfolio. The predictive power of a variety of explanatory variables is examined extensively in the context of both univariate and multivariate regressions. Throughout our empirical analysis, the importance of joint inference on more than one predictive variable is revealed. We present important cases where a set of predictive variables is jointly significant for stock returns whereas each variable in the set has insignificant predictive value.

Chapter 5 concludes the discussion of this thesis.

CHAPTER 2

The Finite-Sample Effects of VAR Dimensions on OLS Bias, OLS Variance, and Minimum MSE Estimators

2.1. Introduction

Vector autoregressions have been extensively studied in econometrics and continue to be one of the most frequently used tools in time series analysis. However, little is currently known about the properties of parameter estimators when applied to finite samples of data, and especially in nonstationary frameworks. In particular, the form and extent of estimator bias and variance have not yet been fully investigated. In a paper that is central to this issue, Abadir, Hadri and Tzavalis (1999) (AHT) study nonstationary multivariate autoregressive series, and derive an approximate expression for the mean bias of the ordinary least squares estimator of the matrix of autoregressive parameters, in terms of the sample size T and VAR dimension k. They consider estimation of a correctly-parameterized first-order vector autoregression (a VAR(1)), with no constant or trend, given that the data generating process is a k-dimensional Gaussian random walk. Using Monte Carlo simulation, they show that their "analytic approximation" provides a good representation of bias in finite samples, and for small k (AHT, Table I).

The purposes of this Chapter are twofold. Firstly, we extend the results given by AHT in a number of directions. In broadening the scope of AHT, we assess over-parameterization of the estimated VAR *model*, by including a constant, and a constant and deterministic trend. This creates additional bias problems, as was suggested by simulation results for the univariate case in Abadir and Hadri (2000, p. 97) and Tanizaki (2000, Table 1). We also assess the effects of introducing $p-1$ excess lags into the estimated model. We use Monte Carlo methods to simulate small sample bias, and then fit a series of response surfaces using weighted nonlinear least squares. Well-specified and parsimonious response surfaces are chosen following diagnostic testing, and are shown to perform very well in out-of-sample prediction. In the correctly-parameterized setting, the prediction error of our response surface is substantially less than that of the AHT form, across the parameter space under investigation.

Secondly, we focus attention on the variance and MSE of the least squares estimator, and generalize the heuristic univariate variance approximation of Abadir (1995) to rigorous response surfaces. We develop response surfaces for variance, and show that multiplying the OLS estimator by a scalar correction factor achieves minimum MSE and removes most of the bias, at the expense of a small increase in estimator variance.[1] To our knowledge, no other finite-sample approximations (analytic or otherwise), and few simulations, were previously available for bias in

[1]See Hendry and Krolzig (2005, section 4) for a similar form of bias correction, after computer-automated model selection.

the multivariate over-parameterized cases, or for excess lags, or for variance in the multivariate setting.

The Chapter is organized as follows. Section 2.2 introduces the possibly over-parameterized VAR model and briefly reviews existing finite-sample results. Section 2.3 outlines the response surface methodology, presents the experimental design, and proposes response surfaces for multivariate bias and variance, based upon an extensive series of Monte Carlo experiments. Section 2.4 concludes the Chapter. We represent vector (and scalar) and matrix quantities as a and A respectively. Special vectors and matrices include the $k \times 1$ zero vector 0_k and the $k \times k$ identity matrix I_k.

2.2. Models and background

Let $\{x_t\}_1^T$ be a $k \times 1$ discrete time series that follows a purely nonstationary VAR(1), where T is the sample size, the innovations are independently and identically distributed with distribution D, and Ω is positive-definite:

$$x_t = x_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \text{i.i.d.D}\left(0_k, \Omega\right), \quad t = 1, 2, \ldots, T. \tag{2.1}$$

We examine the finite-sample bias, variance and MSE of the least squares estimator of (2.1), for each of the following estimated VAR(p) models:

$$
\begin{aligned}
\text{Model A}: \quad x_t &= \widehat{\Phi} x_{t-1} + \sum_{j=1}^{p-1} \widehat{\Gamma}_j \Delta x_{t-j} + \widehat{\varepsilon}_t, \\
\text{Model B}: \quad x_t &= \overline{\mu} + \overline{\Phi} x_{t-1} + \sum_{j=1}^{p-1} \overline{\Gamma}_j \Delta x_{t-j} + \overline{\varepsilon}_t, \\
\text{Model C}: \quad x_t &= \widetilde{\mu} + \widetilde{\delta} t + \widetilde{\Phi} x_{t-1} + \sum_{j=1}^{p-1} \widetilde{\Gamma}_j \Delta x_{t-j} + \widetilde{\varepsilon}_t,
\end{aligned}
$$

where Δ is the backward-difference operator, and over-parameterization arises through inclusion of a constant (Model B), a constant and time trend (Model C), and when there are multiple lags, with $p > 1$ (Models A, B, and C).[2] There are no elements in the summations if $p = 1$. Zero initial values are chosen for simplicity ($x_{-j} = 0_k$, $j = 0, 1, \ldots, p-1$), and to avoid the problems of bias nonmonotonicity that can potentially arise when non-zero initial values are considered.[3]

PROPOSITION 2.1: *The bias matrix* $B = E(\widehat{\Phi}) - I_k$ *is scalar, and bias is invariant to* Ω*, for Models A, B, and C, if the error distribution* D *is symmetric, and* Ω

[2]We are very grateful to the referees of the Journal of Econometrics, who suggested that we generalize our original models.

[3]The correctly-parameterized univariate Model A, with $k = p = 1$, was examined by Abadir and Hadri (2000), given a (nearly) nonstationary data generating process, and non-zero initial values. They show, using numerical integration, that the bias of $\widehat{\phi}$ can be increasing in sample size T, due to the effect of $|x_0|$. This nonmonotonicity disappears under estimation of univariate Models B and C, at the expense of higher bias. A small simulation study of (1) and Model A by Lawford (2001), with $k \leq 6$, $p = 1$ and $x_0 \neq 0_k$, leads to the interesting conjecture that bias nonmonotonicity also disappears when $k > 1$.

is positive-definite. Furthermore, the variances of each of the diagonal elements of $\widehat{\Phi}$ are identical, and variance is invariant to Ω, for Models A, B, and C, if D is symmetric, and Ω is both positive-definite and diagonal.

Proof of the above Proposition is given in Section 2.5. Abadir (1993) uses some results on moment generating functions to derive a high-order closed form (integral-free) analytical approximation to the *univariate* finite-sample bias of $\widehat{\phi}$ given Model A, $k = p = 1$, and with $|\phi| = 1$. The final expression is based upon parabolic cylinder functions, and is computationally very efficient. Abadir further shows that bias may be described more simply in terms of exponential functions in polynomials of T^{-1}, and develops the following heuristic approximation:

$$b^{\text{UNIV}} \approx -1.7814\, T^{-1} \exp\left(-2.6138\, T^{-1}\right), \tag{2.2}$$

where -1.7814 is the expected value of the limiting distribution of $T(\widehat{\phi} - 1)$, e.g. see Le Breton and Pham (1989, p. 562).[4] Heuristic fits such as (2.2) have been used elsewhere in the literature, e.g. Dickey and Fuller (1981, p. 1064), and we distinguish here between these approximations and the rigorous response surface approach that is used in this Chapter. Despite the fact that only 5 datapoints are used in the derivation of (2.2), it is accurate in-sample to 5 decimal places for bias, and is more accurate than the special function expression (see Abadir, 1993,

[4]This constant can be calculated conveniently by using the expression $1 - \frac{1}{2}\int_0^\infty u\,(\cosh u)^{-1/2}\,\mathrm{d}u = 1 - 2\sqrt{2}\;{}_3F_2\,(1/4, 1/4, 1/2; 5/4, 5/4; -1) \approx -1.7814$, where ${}_3F_2$ is a hypergeometric function.

Table 1). We found that (2.2) also performs very well out-of-sample, at least to 1 decimal place of $-100\times$bias. Other studies that examine the exact moments of OLS in univariate autoregressive models, with a variety of disturbances, include Evans and Savin (1981), Nankervis and Savin (1988), Tsui and Ali (1994), and Vinod and Shenton (1996); see also Maeshiro (1999) and Tanizaki (2000), and references therein.

In the multivariate setting, AHT consider Model A, $k \geq 1$, $p = 1$, and prove that B is exactly a *scalar matrix*, i.e. diagonal with equal diagonal elements: $B = \mathrm{diag}(b, \ldots, b)$, and that B is invariant to Ω, given only a symmetric error distribution. Furthermore, they develop a simple quantitative approximation to *multivariate* finite-sample bias (especially AHT, p. 166, and Abadir, 1995, p. 264):

$$B^{\mathrm{AHT}} \approx b^{\mathrm{UNIV}} k I_k \equiv b^{\,\mathrm{AHT}} I_k. \tag{2.3}$$

It is clear that bias is approximately proportional to the dimension of the VAR, even when Ω is diagonal. To facilitate discussion of cointegrating relations, AHT formulate their model as $\triangle x_t = \Psi x_{t-1} + \varepsilon_t$, where $\Psi \equiv \Phi - I_k$. Since the bias of $\widehat{\Psi}$ is equivalent to the bias of $\widehat{\Phi}$, our results may be compared directly to those in AHT, for $p = 1$, and no deterministics.

Abadir (1995, p. 265) uses the univariate Model A ($p = 1$) variance definition $v = 2T^{-2}\mathrm{SD}^2$, with values for standard deviation "SD" of normalized $\widehat{\phi}$ taken from Evans and Savin (1981, Table III), and performs a similar heuristic process to that used in derivation of (2.2) for bias. This gives a variance approximation for $k = p = 1$:

$$v^{\mathrm{UNIV}} \approx 10.1124\,T^{-2} \exp\left(-5.4462\,T^{-1} + 14.519\,T^{-2}\right), \tag{2.4}$$

which is shown to be accurate to at least 7 decimal places in small samples. Since the bias and variance of each of the diagonal elements of $\widehat{\Phi}$ are respectively identical, we may use $\mathrm{MSE}(\widehat{\phi}) = b^2 + v$ directly, to compute the MSE.

In the following Section, we present the Monte Carlo experimental design, develop very accurate response surface approximations to multivariate bias and variance, and consider a simple correction for the OLS estimator to have minimum MSE.

2.3. Structure of Monte Carlo analysis

2.3.1. Response surfaces

Response surfaces are numerical-analytical approximations, which can be very useful when summarizing and interpreting the small sample behaviour of tests and estimators. They have been applied to a variety of econometric problems by, *inter alia*, Engle, Hendry and Trumble (1985), Campos (1986), Ericsson (1991), MacKinnon (1994, 1996), Cheung and Lai (1995), MacKinnon, Haug and Michelis (1999) and Ericsson and MacKinnon (2002). The response surface technique aims to summarize the behaviour of the statistic of interest at all points in the admissible parameter space, i.e. for whole families of DGP's; and in a more sophisticated manner than that offered by simple heuristic approximations. The following outline of the methodology draws upon Hendry (1984) and Davidson and MacKinnon (1993, pp. 755-763).

The quantity of interest τ is a function of the sample size T and the vector of variables θ that appear in the DGP. The relationship is modelled as a functional form $\Psi(T, \theta; \omega)$, where ω is a vector of parameters to be estimated, and $\Psi(\cdot)$ is chosen by the investigator. Estimated values for the dependent variable, $\widetilde{\tau}_i$, are generated using a set of N Monte Carlo experiments. The i^{th} experiment is associated with an estimated standard error $\widetilde{\sigma}(\widetilde{\tau}_i)$, where $\widetilde{\tau}_i$ is approximately distributed as $\mathrm{N}\big(\Psi(T, \theta; \omega), \widetilde{\sigma}^2(\widetilde{\tau}_i)\big)$ if the number of replications per experiment (M) is large. Given that each of the experiments uses different sets of random numbers, we may then implement generalized least squares (with a fully specified covariance matrix) and estimate

$$\frac{\widetilde{\tau}_i}{\widetilde{\sigma}(\widetilde{\tau}_i)} = \frac{\Psi(T, \theta; \omega)}{\widetilde{\sigma}(\widetilde{\tau}_i)} + \varepsilon_i; \quad \varepsilon_i \sim \mathrm{IN}(0, 1); \quad i = 1, \ldots, N, \tag{2.5}$$

using ordinary or nonlinear least squares, depending upon the form chosen for $\Psi(T, \theta; \omega)$. Division by $\widetilde{\sigma}(\widetilde{\tau}_i)$ in (2.5) corrects for heteroscedasticity.

There are a number of potential difficulties associated with the approach. Firstly, precise estimates are needed if $\Psi(T, \theta; \omega)$ is to be accurately specified.

Since a large number (N) of datapoints is also needed – and in practice this seems to be rather more important than having extremely accurate datapoints (although M must be reasonably large) – the method tends to be computationally intensive. Secondly, the functional form of $\Psi(T, \theta; \omega)$ is generally not known *a priori*. Thus, estimation of correctly specified response surfaces becomes very difficult indeed as the number of parameters in $\Psi(T, \theta; \omega)$ increases. Generally, $\Psi(T, \theta; \omega)$ should be formulated in line with known analytical results (as, e.g. we have here in (2.2) and (2.3)).

Thirdly, Monte Carlo studies can be subject to *specificity* of the results, i.e. while the estimated response surface may fit well in-sample, there is no guarantee that accuracy will be achieved over the entire domain of approximation (Hendry, 1984). To avoid this, T and θ should be chosen to span an "interesting" part of the parameter space, (e.g. more detail may be given to sample sizes that are typical in economic applications), and the estimated response surface subjected to a battery of standard diagnostic tests. One useful check of the suitability of the response surface specification is that we would expect a unit error variance, after the heteroscedasticity transformation. Inevitably, some (and often a great deal of) experimentation will be required before correctly specified and parsimonious equations can be selected. The accuracy of the approximation should then be examined using out-of-sample parameter values, i.e. points that are not used in estimation of the response surface. This provides a rigorous test of the accuracy of the method and, if the response surface is correctly specified, will enable the statistic of interest to be approximated at various parameter points without the need to carry out another simulation. It is important to report the parameter values used in the simulation experiments; and extreme caution should be exercised when inferring any findings to more general situations than those defined by the DGP and the specific parameter environment (see especially Maasoumi and Phillips, 1982).

2.3.2. Monte Carlo design and simulation

The data generating process and models were introduced in (2.1) and Models A, B, and C. We adopt a minimal complete factorial design, which covers all triples (T, k, p) from:

$$T \in \{20, 21, \ldots, 30, 35, \ldots, 80, 90, 100, 150, 200\}, \; k \in \{1, 2, 3, 4\}, \; p \in \{1, 2, 3, 4\}, \tag{2.6}$$

giving $N = 400$ datapoints. The sample sizes that we have chosen are representative of those that are commonly used in practice, and our design includes small k and p, so that the effects of changes in VAR dimension and model lag can be explored. From Proposition 1, and with no loss of generality, we set $\varepsilon_t \sim$ i.i.d.N$(0_k, I_k)$ in the simulations. We calculate the OLS estimate for each combination of (T, k, p) in the parameter space, from which we directly derive the bias. Since B is a scalar matrix, we may estimate the scalar b by averaging over the estimated diagonal elements of B. This results in a further increase in accuracy as

k increases. We simulate variance v similarly.[5] The period of our pseudo-random number procedure is much larger than the total random number requirement. All simulations were performed most recently on Pentium 4 machines, with 2.5GHz processors and 512MB of RAM, running GAUSS and/or Python under Microsoft Windows XP.

Where possible, our numerical results were checked with partial exact and approximate results in the literature. These include MacKinnon and Smith (1998, Figure 1), who plot bias functions under Model B ($k = p = 1$), and Pere (2000, Table 3), who reports values that correspond to variances in the same model, in his study of adjusted profile likelihood. Evans and Savin (1981, Table 3) give bias and standard deviation for $2^{-1/2}T(\widehat{\phi} - 1)$ under Model A ($k = p = 1$), which agree closely (3 to 5 decimal places) with our simulation results. Roy and Fuller (2001, Tables 1 and 6) report bias and MSE for $T = 100$, under univariate Models B and C, for $p = 1$.

2.3.3. Post-simulation analysis

We regressed the Monte Carlo estimates of bias and variance under Models A, B, and C, on functions of sample size, VAR dimension and lag order, to reflect

[5] We experimented with a pseudo-antithetic variate technique, based upon Abadir and Paruolo's (2009) univariate "AV4", and were able to increase the speed of the bias simulations by roughly 50%, for a given precision [Model A, $p = 1$]. While conventional antithetics are not generally applicable to the nonstationary setting, the pseudo-antithetic is not valid either for some of the models considered above, and is therefore not used in this paper.

the dependence of b and v upon these parameters, and on the degree of over-parameterization. Following extensive experimentation, and motivated by (2.2), we fit the following nonlinear bias response surface for each of the models:[6]

$$(s_i^b)^{-1}\, b\,(T_i, k_i, p_i) = (s_i^b)^{-1}\,(\beta_1 + \beta_2\, k_i)\, T_i^{-1} \exp\left[(\beta_3 + \beta_4\, k_i + \beta_5\, k_i p_i + \beta_6\, k_i^{p_i})\, T_i^{-1}\right] + u_i. \tag{2.7}$$

The dependent variable $b\,(T_i, k_i, p_i)$ is the simulated finite-sample bias for sample size T_i, VAR dimension k_i, and lag order p_i, which take values from (2.6), and u_i is an error term. We correct for Monte Carlo sampling heteroscedasticity using the term s_i^b, which is the simulated sampling error standard deviation of bias over replications (see Doornik and Hendry, 2007, Chapter 15, for details). We denote the fitted values of the estimated response surface by b^{RS}, and estimated coefficients are reported in Table 1. Convergence of the weighted nonlinear least squares routine was very fast, and required few iterations. Selection criteria included small residual variance and good in-sample fit, parsimony, and satisfactory diagnostic performance. The response surface fits are extremely good in-sample, and the Jarque-Bera statistic for normality is small. The signs of all estimated coefficients apart from the constant β_1 remain the same across the different models. Note

[6]Some early motivation for numerical refinement of (2.3), for Model A, with $p = 1$, came from consideration of low-order partial derivatives of b^{AHT} . Straightforward algebra gives (for $T \geq 1$) $b^{\mathrm{AHT}} < 0$, $\partial b^{\mathrm{AHT}}/\partial k < 0$, $\partial^2 b^{\mathrm{AHT}}/\partial k^2 = 0$, (for $T \geq 3$) $\partial b^{\mathrm{AHT}}/\partial T > 0$, $\partial^2 b^{\mathrm{AHT}}/\partial k \partial T > 0$, (for $T \geq 5$) $\partial^2 b^{\mathrm{AHT}}/\partial T^2 < 0$. Upon comparing these theoretical partials with approximate numerical partial derivatives from simulated data, it is found that each holds, except for $\partial^2 b/\partial k^2 = 0$ (simulations suggest that $\partial^2 b/\partial k^2 > 0$, for T not too large). This finding suggested that improvements were possible over (2.3), and especially that k entered the formula in a more complicated manner than in (2.3).

that the asymptotic bias $T_i b$ (as $T_i \to \infty$) is a linear function of k_i alone, which agrees with numerical observations, and that $\beta_1 + \beta_2 k_i$ can be interpreted as the asymptotic component of bias, with the exponential representing the (analytically intractable) finite-sample "adjustment", which depends on k_i and p_i (and T_i).

We recalculate Table I in AHT as Table 2 in this Chapter, with increased accuracy, with additional results reported for $T = 400, 800$ and $k = 6, 7, 8$, and correcting for a typo in AHT Table I: $(T, k) = (25, 5)$. It is convenient to interpret the scaled bias values as percentages of the true parameter value, e.g. in Model A, given $(T, k) = (25, 8)$, and $p = 1$, the absolute bias of each of the estimated parameters on the diagonal of $\widehat{\Phi}$ is 46.7% of the true value (unity). Clearly, absolute bias is strictly increasing in k and decreasing in T. As T increases, bias goes to zero, as is well-known from asymptotic theory. We see that b^{AHT} gives a good approximation to bias for k small, and especially for $k = 1$, where (2.3) reduces to the excellent heuristic approximation (2.2). However, as k increases, b^{RS} provides much closer approximations to bias, even for T quite large. Out-of-sample points reported in Table 2 for b^{RS} are combinations of $k = 5, 6, 7, 8$, and $T = 400, 800$. While b^{AHT} is only applicable for correctly-parameterized Model A, our response surfaces can be used when $p > 1$, and also when deterministics are included. The out-of-sample fit appears to be excellent for all T, and up to about $k = p = 6$ (as k and p jointly become large, with small T, the term k^p will dominate the bias approximation, and out-of-sample predictions should be used with particular caution). Although the response surfaces are developed with small sample rather than asymptotic considerations in mind, it is interesting to approximate univariate asymptotic bias by setting $k = p = 1$ and letting $T_i \to \infty$ in $T_i b^{\mathrm{RS}}$, from (2.7), which gives $T_i b^{\mathrm{RS}} = \widehat{\beta}_1 + \widehat{\beta}_2$ of approximately -1.7, -5.4 and -10.3 in Models A, B and C respectively.

Kiviet and Phillips (2005, equation (14), and Figure 1) consider univariate Model B, where the data generating process can have a non-zero drift, and write autoregressive bias in terms of "g-functions " $g_0(T)$ and $g_1(T)$, which they calculate using simulations. The function $g_0(T)$ represents least squares bias when there is a zero drift in the data generating process, while $g_1(T)$ appears as the bias increment due to non-zero drift. Our equation (2.7) simplifies (when $k = p = 1$) to $g_0(T) \approx -5.3654\, T^{-1} \exp\left(-2.6513\, T^{-1}\right)$, which provides a convenient means of calculating $g_0(T)$ without further simulations.

Using (2.4) to motivate the choice of functional form, we fit the variance response surface:

$$\begin{aligned}(s_i^v)^{-1} v\left(T_i, k_i, p_i\right) \quad = \quad & (s_i^v)^{-1} \left(\gamma_1 + \gamma_2\, k_i\right) T_i^{-2} \exp[\left(\gamma_3 + \gamma_4\, k_i + \gamma_5\, p_i + \gamma_6\, k_i p_i\right) T_i^{-1} \\ & + \left(\gamma_7\, k_i p_i + \gamma_8\, k_i^{p_i}\right) T_i^{-2}] + u_i, \qquad (2.8)\end{aligned}$$

where $v\left(T_i, k_i, p_i\right)$ is the simulated finite-sample variance, and s_i^v is the simulated sampling error standard deviation of the variance over replications. In estimating (2.8), we did not use datapoints for which $T_i = 20, \ldots, 24$ (and so $N = 320$), since variance becomes very large for such small sample sizes, which makes it very difficult to specify good response surfaces across the full parameter space. Estimated response surfaces v^{RS} are given in Table 3, and are seen to fit very well. The signs of each of the estimated coefficients, except for γ_1, remains the same across the models, the Jarque-Bera statistic is relatively low, and v^{RS} provides a very good approximation in-sample. The out-of-sample variance approximation should be used with caution as k and p jointly exceed about 5 or 6, with small T, again due to the effect of the term k^p. We note that no variance approximations were previously available for over-parameterized models, excess lags, or even for $k > 1$. Similarly to the bias response surfaces, the asymptotic variance $T_i^2 v$ (as $T_i \to \infty$) is a linear function of k_i alone, and $\gamma_1 + \gamma_2 k_i$ can be interpreted as the asymptotic component of variance, with the exponential term again representing the finite-sample "adjustment", which depends upon k_i and p_i (and T_i). The dependencies of bias and variance on T, k, and p are depicted in Figures 1 and 2, which plot scaled response surfaces $-100 \times b^{\mathrm{RS}}$ and $10,000 \times v^{\mathrm{RS}}$, against T, for Models A, B, and C, with $k = 1, 2$ and $p = 1, 2$.

Bias and variance are not the only criteria to be used in comparison of time series estimates, and the mean squared error, $\mathrm{MSE}(\widehat{\phi}) = b^2 + v$, is often of interest. For univariate Model A ($p = 1$), Abadir (1995) defines λ as a correction factor such that $\mathrm{MSE}(\lambda\widehat{\phi})$ is minimized, and b^m and v^m as the bias and variance of the corrected OLS estimator $\lambda\widehat{\phi}$, with:

$$\lambda = \frac{1+b}{v + (1+b)^2}, \quad b^m = \frac{-v}{v + (1+b)^2}, \quad v^m = \lambda^2 v, \tag{2.9}$$

when $\phi = 1$. Equations (2.7) and (2.8) may be now combined to give an approximation to MSE, and by substitution of response surface values for bias and variance into (2.9), we are able to calculate λ for various T, k, and p. As an illustration, correction factors are reported in Table 4, for $p = 1$ and Model A, which displays qualitatively similar results to those in Abadir (1995, Tables 2 and 3). It is clear that OLS ($\lambda = 1$) does not achieve minimum MSE. It is also shown that the corrected OLS is almost unbiased, unlike OLS. From Table 4, λ increases monotonically with k and decreases monotonically with T (asymptotically, the OLS achieves minimum MSE). The correction can be particularly large for small T, e.g. $(T, k) = (25, 5)$ implies a correction of 32%. The corrected estimator is much less biased than the OLS, and b^m tends to zero more rapidly than b. However, this reduction in bias comes at the expense of a small increase in the variance of the corrected estimator, v^m. It is seen that b^2 forms a much larger proportion of MSE than variance for $k \geq 3$, although this does not hold following the minimum MSE correction; and that MSE efficiency is generally decreasing in T and k.

2.4. Concluding comments

We have performed an extensive set of Monte Carlo experiments on the bias and variance of the OLS of the autoregressive parameters, given a data generating process that is a purely nonstationary VAR(1), where the estimated model is a possibly over-parameterized VAR(p), for small sample sizes, and various VAR dimensions and lag lengths. Although the univariate framework has been the subject of much research, a comprehensive multivariate simulation study has not previously been performed. We estimate parsimonious and computationally convenient response surfaces for bias and variance, that are much more accurate and more general than existing approximations. In this way, we improve numerically upon existing finite-sample analytical bias results, and extend them to $p > 1$ and deterministics, and also extend existing finite-sample variance results to $k > 1$, $p > 1$, and to deterministics. Finally, we investigate the correction factors required for the OLS to achieve minimum MSE and show that this correction can significantly reduce bias, at the expense of a small increase in estimator variance. Our results may provide guidelines for applied researchers as to the behaviour of VAR models, given that relatively short samples and nonstationary data are often relevant in empirical work.

Our work complements important asymptotic treatments by Phillips (1987a) in the univariate framework, and Park and Phillips (1988, 1989), Phillips (1987b), and Tsay and Tiao (1990) in the multivariate setting. Our results may also be useful when studying the derivation of exact formulae (for instance, in conjunction with work by Abadir and Larsson, 1996, 2001, who derive the exact finite-sample moment generating function of the quadratic forms that create the basis for the sufficient statistic in a discrete Gaussian vector autoregression). Exact analytical bias expressions may involve multiple infinite series of matrix-argument hypergeometric functions (generalizing, e.g. Abadir, 1993). When such series arise in other areas of econometrics, they are generally complicated and may be difficult to implement for numerical evaluation. We may, therefore, need to rely upon approximations in practice, even when the exact formulae are available.

2.5. Technical appendix and proofs

Proof of Proposition 2.1. Let ε_t be i.i.d. $N(0_k, \Omega)$

$$x_t = Rx_{t-1} + \varepsilon_t = \sum_{j=1}^{t} R^{t-j}\varepsilon_j \tag{2.10}$$

with $R = I_k$, and setting $x_0 = 0$.

We are going to deal with the most general case, which refers to the estimation of:

$$x_t = \widetilde{\mu} + \widetilde{\delta}t + \widetilde{\Phi}x_{t-1} + \sum_{j=1}^{p-1} \widetilde{\Gamma}_j \Delta x_{t-j} + \varepsilon_t.$$

We need to show that the bias of matrix Φ is scalar. A sufficient condition for this, using the following model

$$x_t = \mu + \delta t + \sum_{i=1}^{p} A_i x_{t-j} + \varepsilon_t, \tag{2.11}$$

is that the bias of any matrix A_i (for $i = 1, ..., p$) is a scalar matrix since

$$\Phi = \sum_{i=1}^{p} A_i. \tag{2.12}$$

Let

$$\begin{aligned}
\Omega &= LL', \\
\mathcal{L} &= I_{(p+2)} \otimes L \\
W_t &= \begin{pmatrix} x_t \\ x_{t-1} \\ \vdots \\ x_{t-p+1} \\ 1_k \\ (t+1)\,1_k \end{pmatrix}, U_t = \begin{pmatrix} \varepsilon_t \\ 0_{(p+1)k\times 1} \end{pmatrix},
\end{aligned}$$

where 1_k is a $k \times 1$ vector with each element being 1. and Then (2.11) can be represented as:

$$W_t = \mathcal{A}W_{t-1} + U_t, \tag{2.13}$$

where

$$\mathcal{A}_{(p+2)k\times(p+2)k} = \begin{pmatrix} A_1 & A_2 & \cdots & A_{p-1} & A_p & \mathcal{C} & \mathcal{D} \\ I_k & 0_{k\times k} & & 0_{k\times k} & 0_{k\times k} & 0_{k\times k} & 0_{k\times k} \\ 0_{k\times k} & I_k & & 0_{k\times k} & 0_{k\times k} & \vdots & \vdots \\ \vdots & \vdots & \ddots & \vdots & \vdots & & \\ 0_{k\times k} & 0_{k\times k} & \cdots & I_k & 0_{k\times k} & 0_{k\times k} & 0_{k\times k} \\ 0_{k\times k} & 0_{k\times k} & & 0_{k\times k} & 0_{k\times k} & I_k & 0_{k\times k} \\ 0_{k\times k} & 0_{k\times k} & & 0_{k\times k} & 0_{k\times k} & I_k & I_k \end{pmatrix},$$

$$\mathcal{C} = \begin{pmatrix} \mu_1 & 0 & \cdots & 0 \\ 0 & \mu_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mu_k \end{pmatrix}, \mathcal{D} = \begin{pmatrix} \delta_1 & 0 & \cdots & 0 \\ 0 & \delta_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \delta_k \end{pmatrix}$$

and

$$U_t \sim N\left(0_{(p+2)k\times k}, \begin{pmatrix} \Omega & 0_{k\times(p+1)k} \\ 0_{(p+1)k\times k} & 0_{(p+1)k\times(p+1)k} \end{pmatrix}\right)$$

Now define

$$\mathcal{Z}_t = \mathcal{L}^{-1} W_t, \tag{2.14}$$

$$\mathcal{U}_t = \mathcal{L}^{-1} U_t. \tag{2.15}$$

At this point it is important to list some results:

$$
\begin{aligned}
\mathcal{L}\mathcal{L}' &= \left(I_{(p+2)} \otimes L\right)\left(I_{(p+2)} \otimes L\right)' = I_{(p+2)} \otimes LL' = I_{(p+2)} \otimes \Omega \\
\mathcal{L}^{-1} &= \left(I_{(p+2)} \otimes L\right)^{-1} = I_{(p+2)} \otimes L^{-1} \\
\mathcal{L}^{-1} W_t &= \left(I_{(p+2)} \otimes L^{-1}\right) \begin{pmatrix} x_t \\ x_{t-1} \\ \vdots \\ x_{t-p+1} \\ 1_k \\ (t+1)\, 1_k \end{pmatrix} = \begin{pmatrix} L^{-1} x_t \\ L^{-1} x_{t-1} \\ \vdots \\ L^{-1} x_{t-p+1} \\ L^{-1} 1_k \\ L^{-1} (t+1)\, 1_k \end{pmatrix} \\
\mathcal{U}_t &= \mathcal{L}^{-1} U_t = \left(I_{(p+2)} \otimes L^{-1}\right) \begin{pmatrix} \varepsilon_t \\ 0_{(p+1)k \times 1} \end{pmatrix} = \begin{pmatrix} L^{-1} \varepsilon_t \\ 0_{(p+1)k \times 1} \end{pmatrix}.
\end{aligned}
$$

Using the Cholesky decomposition

$$
\mathcal{L}^{-1} E\left(\widehat{\mathcal{A}} - \mathcal{A}\right)\mathcal{L} = E\left(\sum_{t=1}^{n} \mathcal{U}_t \mathcal{Z}_{t-1}'\right)\left(\sum_{t=1}^{n} \mathcal{Z}_{t-1} \mathcal{Z}_{t-1}'\right)^{-1}. \tag{2.16}
$$

The value of $\mathcal{A}$ under the DGP is

$$\mathcal{A} = \begin{pmatrix} I_k & 0_{k\times k} & \cdots & 0_{k\times k} & 0_{k\times k} & 0_{k\times k} & 0_{k\times k} \\ I_k & 0_{k\times k} & \cdots & 0_{k\times k} & 0_{k\times k} & 0_{k\times k} & 0_{k\times k} \\ 0_{k\times k} & I_k & \cdots & 0_{k\times k} & 0_{k\times k} & 0_{k\times k} & 0_{k\times k} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0_{k\times k} & 0_{k\times k} & \cdots & I_k & 0_{k\times k} & 0_{k\times k} & 0_{k\times k} \\ 0_{k\times k} & 0_{k\times k} & \cdots & 0_{k\times k} & 0_{k\times k} & I_k & 0_{k\times k} \\ 0_{k\times k} & 0_{k\times k} & \cdots & 0_{k\times k} & 0_{k\times k} & I_k & I_k \end{pmatrix}.$$

Pre-multiplying both parts of (2.13) by $\mathcal{L}^{-1}$:

$$\mathcal{L}^{-1}W_t = \mathcal{L}^{-1}\mathcal{A}W_{t-1} + \mathcal{L}^{-1}U_t. \tag{2.17}$$

A crucial property for the proof is the fact that matrices $\mathcal{L}^{-1}$ and $\mathcal{A}$ commute:

$$\mathcal{L}^{-1}\mathcal{A} = \mathcal{A}\mathcal{L}^{-1}. \tag{2.18}$$

Using (2.18) in (2.17) and definitions (2.14) and (2.15) we get

$$\begin{aligned} \mathcal{Z}_t &= \mathcal{L}^{-1}\mathcal{A}W_{t-1} + \mathcal{U}_t = \mathcal{A}\mathcal{L}^{-1}W_{t-1} + \mathcal{U}_t \\ &= \mathcal{A}\mathcal{Z}_{t-1} + \mathcal{U}_t, \end{aligned} \tag{2.19}$$

with

$$\begin{aligned}\mathcal{U}_t &\sim N(0,\Psi),\\ \Psi &= \begin{pmatrix} I_k & 0_{k\times(p+1)k} \\ 0_{(p+1)k\times k} & 0_{(p+1)k\times(p+1)k}\end{pmatrix}.\end{aligned}$$

Define matrix G_q such that

$$\begin{aligned}G_1 &= \begin{pmatrix} -1 & 0_{1\times(k-1)} \\ 0_{(k-1)\times 1} & I_{k-1}\end{pmatrix}\\ G_p &= \operatorname{diag}(I_{p-1},-1,I_{k-p}) \quad p\geq 2.\end{aligned}$$

Note that $G_q^{-1} = G_q$ for all $q \in \{1,..,k\}$.

Define

$$\Xi_q = I_{(p+2)} \otimes G_q,$$

and let

$$\widetilde{\mathcal{U}}_t = \Xi_q \mathcal{U}_t, \tag{2.20}$$

$$\widetilde{\mathcal{Z}}_t = \Xi_q \mathcal{Z}_t. \tag{2.21}$$

Again we use the commutation property that holds for matrices $\mathcal{A}$ and Ξ_q :

$$\Xi_q \mathcal{A} = \mathcal{A}\Xi_q \tag{2.22}$$

Left multiplying both sides of (2.19) by Ξ_q, and using definitions (2.20) and (2.21) and commutation property in (2.22) we obtain

$$\begin{aligned}\Xi_q \mathcal{Z}_t &= \Xi_q \mathcal{A} \mathcal{Z}_{t-1} + \Xi_q \mathcal{U}_t \Rightarrow \widetilde{\mathcal{Z}}_t = \mathcal{A} \Xi_q \mathcal{Z}_{t-1} + \widetilde{\mathcal{U}}_t \\ &\Rightarrow \widetilde{\mathcal{Z}}_t = \mathcal{A} \widetilde{\mathcal{Z}}_{t-1} + \widetilde{\mathcal{U}}_t \qquad (2.23)\end{aligned}$$

with

$$\widetilde{\mathcal{U}}_t \sim N(0, \Psi).$$

Now let

$$\begin{aligned} B &= E\left(\sum_{t=1}^{n} U_t W'_{t-1}\right)\left(\sum_{t=1}^{n} W_{t-1} W'_{t-1}\right)^{-1} \\ \mathcal{B} &= E\left(\sum_{t=1}^{n} \mathcal{U}_t \mathcal{Z}'_{t-1}\right)\left(\sum_{t=1}^{n} \mathcal{Z}_{t-1} \mathcal{Z}'_{t-1}\right)^{-1} \qquad (2.24) \\ \widetilde{\mathcal{B}} &= E\left(\sum_{t=1}^{n} \widetilde{\mathcal{U}}_t \widetilde{\mathcal{Z}}'_{t-1}\right)\left(\sum_{t=1}^{n} \widetilde{\mathcal{Z}}_{t-1} \widetilde{\mathcal{Z}}'_{t-1}\right)^{-1} \end{aligned}$$

and using the definition of $\widetilde{\mathcal{U}}_t$ and $\widetilde{\mathcal{Z}}_t$ we obtain

$$\mathcal{L}^{-1} E\left(\widehat{\mathcal{A}} - \mathcal{A}\right)\mathcal{L} = \mathcal{B}$$

$$\begin{aligned} \mathcal{B} &= \mathcal{L}^{-1} B \mathcal{L} \\ \tilde{\mathcal{B}} &= \Xi_q \mathcal{B} \Xi_q. \qquad (2.25) \end{aligned}$$

In what follows, we prove that

$$\tilde{\mathcal{B}} = \mathcal{B} \text{ for all } q \in \{1, .., k\}. \tag{2.26}$$

From (2.19) we have

$$\mathcal{Z}_t = \sum_{j=1}^{t} \mathcal{A}^{t-j} \mathcal{U}_j \Rightarrow \mathcal{Z}_{t-1} = \sum_{j=1}^{t-1} \mathcal{A}^{t-1-j} \mathcal{U}_j, \tag{2.27}$$

and from (2.23)

$$\widetilde{\mathcal{Z}}_{t-1} = \sum_{j=1}^{t-1} \mathcal{A}^{t-1-j} \widetilde{\mathcal{U}}_j. \tag{2.28}$$

Using (2.27) and (2.28) and the independence of U_t we get that

$$E\left[\mathcal{U}_n \mathcal{Z}_{n-1} \left(\sum_{t=1}^{n} \mathcal{Z}_{t-1} \mathcal{Z}'_{t-1}\right)^{-1}\right] = 0.$$

Combining the above we conclude that

$$\mathcal{B} = h\left(\mathcal{U}_{n-1}, ..., \mathcal{U}_1\right) \tag{2.29}$$

$$\tilde{\mathcal{B}} = h\left(\widetilde{\mathcal{U}}_{n-1}, ..., \widetilde{\mathcal{U}}_1\right). \tag{2.30}$$

for some function h. Denoting by $f\left(\mathcal{U}_{n-1}, ..., \mathcal{U}_1\right)$ the joint density of $\left(\mathcal{U}_{n-1}, ..., \mathcal{U}_1\right)$, independence gives

$$f\left(\mathcal{U}_{n-1}, ..., \mathcal{U}_1\right) = f\left(\mathcal{U}_{n-1}\right) ... f\left(\mathcal{U}_1\right). \tag{2.31}$$

Noting that for any diagonal matrix M we have the property $\Xi_q M \Xi_q = M$,

$$\widetilde{\mathcal{U}}_t = \Xi_q \mathcal{U}_t \sim N\left(0, \Xi_q \Psi \Xi_q'\right) = N\left(0, \Psi\right)$$

showing that $f\left(\widetilde{\mathcal{U}}_t\right) = f\left(\mathcal{U}_t\right)$ for all t. Therefore, (2.31) and independence of the sequence $\left(\widetilde{\mathcal{U}}_t\right)$ give

$$f\left(\mathcal{U}_{n-1}, ..., \mathcal{U}_1\right) = f\left(\widetilde{\mathcal{U}}_{n-1}\right) ... f\left(\widetilde{\mathcal{U}}_1\right) = f\left(\widetilde{\mathcal{U}}_{n-1}, ..., \widetilde{\mathcal{U}}_1\right).$$

Therefore, $\left(\widetilde{\mathcal{U}}_{n-1}, ..., \widetilde{\mathcal{U}}_1\right)$ has the same distribution as $\left(\mathcal{U}_{n-1}, ..., \mathcal{U}_1\right)$, so, for any function h,

$$Eh\left(\widetilde{\mathcal{U}}_{n-1}, ..., \widetilde{\mathcal{U}}_1\right) = Eh\left(\mathcal{U}_{n-1}, ..., \mathcal{U}_1\right).$$

Now (2.29) and (2.30) show (2.26).

Using (2.25) and (2.26) we can conclude that $\mathcal{B}_i$,matrices (defined below) are diagonal. The argument goes as follows: let

$$\mathcal{B} = \mathcal{L}^{-1} E\left(\widehat{\mathcal{A}} - \mathcal{A}\right)\mathcal{L}$$

$$= \begin{pmatrix} L^{-1}E\left(\widehat{A}_1 - A_1\right)L & \cdots & L^{-1}E\left(\widehat{A}_p - A_p\right)L & L^{-1}E\left(\widehat{\mathcal{C}} - \mathcal{C}\right)L & L^{-1}E\left(\widehat{\mathcal{D}} - \mathcal{D}\right)L \\ 0_{k(p+1)\times k} & \cdots & 0_{k(p+1)\times k} & 0_{k(p+1)\times k} & 0_{k(p+1)\times k} \end{pmatrix}$$

$$\equiv \begin{pmatrix} \mathcal{B}_1 & \cdots & \mathcal{B}_p & \mathcal{B}_{p+1} & \mathcal{B}_{p+2} \\ 0_{k(p+1)\times k} & \cdots & 0_{k(p+1)\times k} & 0_{k(p+1)\times k} & 0_{k(p+1)\times k} \end{pmatrix}.$$

Note that $A_1 = I_k, A_2 = A_3 = ... = A_p = \mathcal{C} = \mathcal{D} = 0_{k\times k}$

Define also

$$\mathcal{B}_1 = \begin{pmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1k} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{k1} & \beta_{k2} & \cdots & \beta_{kk} \end{pmatrix} \tag{2.32}$$

Then (2.25) yields

$$\begin{aligned} \tilde{\mathcal{B}} &= \Xi_1 \mathcal{B} \Xi_1 \\ &= \begin{pmatrix} G_1\mathcal{B}_1G_1 & G_1\mathcal{B}_2G_1 & \cdots & G_1\mathcal{B}_pG_1 & G_1\mathcal{B}_{p+1}G_1 & G_1\mathcal{B}_{p+2}G_1 \\ 0_{k(p+1)\times k} & 0_{k(p+1)\times k} & \cdots & 0_{k(p+1)\times k} & 0_{k(p+1)\times k} & 0_{k(p+1)\times k} \end{pmatrix} \end{aligned}$$

and we have that

$$\widetilde{\mathcal{B}}_1 := G_1\mathcal{B}_1G_1 = \begin{pmatrix} \beta_{11} & -\beta_{12} & \cdots & -\beta_{1k} \\ -\beta_{21} & \beta_{22} & \cdots & \beta_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ -\beta_{k1} & \beta_{k2} & \cdots & \beta_{kk} \end{pmatrix}$$

and comparing this to (2.26) and (2.32) we conclude all elements in the first row and first column of $E\left(\widehat{A}_1 - A_1\right)$ apart from β_{11} must be 0:

$$\mathcal{B}_1 = \begin{pmatrix} \beta_{11} & 0 & \cdots & 0 \\ 0 & \beta_{22} & \cdots & \beta_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \beta_{k2} & \cdots & \beta_{kk} \end{pmatrix}. \tag{2.33}$$

Following the same rationale we can understand that matrices $\mathcal{B}_2, \mathcal{B}_3, ..., \mathcal{B}_p$ all have elements in the first row and first column of being 0, apart from the element at position 11.

Using (2.25) on (2.33) we obtain

$$\begin{aligned} \tilde{\mathcal{B}} &= \Xi_2 \mathcal{B} \Xi_2 \\ &= \begin{pmatrix} G_2\mathcal{B}_1G_2 & G_2\mathcal{B}_2G_2 & \cdots & G_2\mathcal{B}_pG_2 & G_2\mathcal{B}_{p+1}G_2 & G_2\mathcal{B}_{p+2}G_2 \\ 0_{k(p+1)\times k} & 0_{k(p+1)\times k} & & 0_{k(p+1)\times k} & 0_{k(p+1)\times k} & 0_{k(p+1)\times k} \end{pmatrix}, \\ &= \begin{pmatrix} B_{11} & 0 & 0 & ... & 0 \\ 0 & B_{22} & -B_{23} & \cdots & -B_{2k} \\ 0 & -B_{32} & B_{33} & \cdots & B_{3k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & -B_{k2} & B_{k3} & \cdots & B_{kk} \end{pmatrix} \end{aligned}$$

and

$$\mathcal{B}_1 = \begin{pmatrix} \beta_{11} & 0 & 0 & ... & 0 \\ 0 & \beta_{22} & -\beta_{23} & ... & -\beta_{2k} \\ 0 & -\beta_{32} & \beta_{33} & ... & \beta_{3k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & -\beta_{k2} & \beta_{k3} & ... & \beta_{kk} \end{pmatrix}$$

and comparing this to (2.26) and (2.33) we conclude all elements in the second row and second column of B apart from B_{22} must be 0 :

$$\mathcal{B}_1 = \begin{pmatrix} \beta_{11} & 0 & 0 & ... & 0 \\ 0 & \beta_{22} & 0 & ... & 0 \\ 0 & 0 & \beta_{33} & ... & \beta_{3k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \beta_{k3} & ... & \beta_{kk} \end{pmatrix}$$

Again the same argument applies to matrices $\mathcal{B}_2, \mathcal{B}_3, ..., \mathcal{B}_p$.

Continuing like this for all $q \in \{1, .., k\}$ we obtain that

$$\mathcal{B}_1 = \operatorname{diag}\left(\beta_{11}, ..., \beta_{kk}\right). \tag{2.34}$$

The same logic applies to $\mathcal{B}_2, \mathcal{B}_3, ..., \mathcal{B}_p$.

To show that each $\mathcal{B}_i$ (for $i = 1, ..., p$) is also scalar, we employ a different linear transformation: Let

$$
\begin{aligned}
\Pi &= \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \\
\Pi_1 &= \begin{pmatrix} \Pi & 0_{2\times(k-2)} \\ 0_{(k-2)\times 2} & I_{k-2} \end{pmatrix} \\
\Pi_r &= \mathrm{diag}\left(I_{r-1}, \Pi, I_{k-r-1}\right) \quad r \geq 2,
\end{aligned}
$$

and

$$\Upsilon_r = I_{(p+2)} \otimes \Pi_r$$

For each $r \in \{1, .., k\}$, let $\overline{\mathcal{Z}}_t = \Upsilon_r \mathcal{Z}_t$ and $\overline{\mathcal{U}}_t = \Upsilon_r \mathcal{U}_t$, and $\overline{\mathcal{U}}_t = \Upsilon_r \mathcal{U}_t$. As before, pre-multiplying both sides of (2.19) by Υ_r gives

$$\overline{\mathcal{Z}}_t = \overline{\mathcal{Z}}_{t-1} + \overline{\mathcal{U}}_t.$$

Letting

$$\bar{\mathcal{B}} = E\left(\sum_{t=1}^{n} \overline{\mathcal{U}}_t \overline{\mathcal{Z}}'_{t-1}\right)\left(\sum_{t=1}^{n} \overline{\mathcal{Z}}_{t-1} \overline{\mathcal{Z}}'_{t-1}\right)^{-1} \tag{2.35}$$

and noting that $\Pi_r = \Pi'_r = \Pi_r^{-1}$, and $\Upsilon_r = \Upsilon'_r = \Upsilon_r^{-1}$ we obtain that

$$\bar{\mathcal{B}} = \Upsilon_r \mathcal{B} \Upsilon_r. \tag{2.36}$$

It is useful here to see the structure of $\bar{\mathcal{B}}$ matrix

$$\begin{aligned}
\bar{\mathcal{B}} &= \Upsilon_r \mathcal{B} \Upsilon_r \\
&= \begin{pmatrix} \Pi_r \mathcal{B}_1 \Pi_r & \Pi_r \mathcal{B}_2 \Pi_r & \cdots & \Pi_r \mathcal{B}_p \Pi_r & \Pi_r \mathcal{B}_{p+1} \Pi_r & \Pi_r \mathcal{B}_{p+2} \Pi_r \\ 0_{k(p+1)\times k} & 0_{k(p+1)\times k} & & 0_{k(p+1)\times k} & 0_{k(p+1)\times k} & 0_{k(p+1)\times k} \end{pmatrix} \\
&\equiv \begin{pmatrix} \overline{\mathcal{B}}_1 & \overline{\mathcal{B}}_2 & \cdots & \overline{\mathcal{B}}_p & \overline{\mathcal{B}}_{p+1} & \overline{\mathcal{B}}_{p+2} \\ 0_{k(p+1)\times k} & 0_{k(p+1)\times k} & & 0_{k(p+1)\times k} & 0_{k(p+1)\times k} & 0_{k(p+1)\times k} \end{pmatrix}
\end{aligned}$$

Since $\overline{\mathcal{U}}_t = \Upsilon_r \mathcal{U}_t \sim N\left(0, \Upsilon_r \Psi \Upsilon_r'\right) = N\left(0, \Psi\right),$ independence of the $(\mathcal{U}_t)$ sequence yields

$$\begin{aligned}
f\left(\overline{\mathcal{U}}_{n-1}, ..., \overline{\mathcal{U}}_1\right) &= f\left(\overline{\mathcal{U}}_{n-1}\right) ... f\left(\overline{\mathcal{U}}_1\right) = f\left(\mathcal{U}_{n-1}\right) ... f\left(\mathcal{U}_1\right) \\
&= f\left(\mathcal{U}_{n-1}, ..., \mathcal{U}_1\right)
\end{aligned}$$

so $\left(\overline{\mathcal{U}}_{n-1}, ..., \overline{\mathcal{U}}_1\right)$ and $\left(\mathcal{U}_{n-1}, ..., \mathcal{U}_1\right)$ have the same distribution and hence

$$Eh\left(\overline{\mathcal{U}}_{n-1}, ..., \overline{\mathcal{U}}_1\right) = Eh\left(\mathcal{U}_{n-1}, ..., \mathcal{U}_1\right)$$

for any function h. The above argument establishes that

$$\bar{\mathcal{B}} = \mathcal{B} \quad \text{for all } r \in \{1, .., k\}\,. \tag{2.37}$$

Now the fact that each $\mathcal{B}_i$ is a scalar matrix follows by (2.36), (2.37) and the fact that Π_r is a permutation matrix: (2.36) and (2.34) give

$$\bar{\mathcal{B}}_1 = \Pi_1 \mathcal{B}_1 \Pi_1 = \operatorname{diag}\left(\beta_{22}, \beta_{11}, \beta_{33}, ..., \beta_{kk}\right).$$

Hence, (2.37) implies that $\beta_{11} = \beta_{22}$. Applying (2.36) with $r = 2$ implies that $\beta_{22} = \beta_{33}$. The same rational applies to the diagonal elements of $\bar{\mathcal{B}}_2, \bar{\mathcal{B}}_3, ..., \bar{\mathcal{B}}_p$. Continuing for all $r \in \{1, .., k\}$ shows that each $\mathcal{B}_i$ is a scalar matrix.

Substituting back to (2.16) we obtain, for some constants $c_1, c_{2,}..., c_{p+2}$,

$$\begin{aligned}
& \mathcal{L}^{-1} E\left(\widehat{\mathcal{A}} - \mathcal{A}\right)\mathcal{L} \\
= \; & \mathcal{B} \\
= \; & \begin{pmatrix} \mathcal{B}_1 & \mathcal{B}_2 & \cdots & \mathcal{B}_p & \mathcal{B}_{p+1} & \mathcal{B}_{p+2} \\ 0_{k(p+1)\times k} & 0_{k(p+1)\times k} & & 0_{k(p+1)\times k} & 0_{k(p+1)\times k} & 0_{k(p+1)\times k} \end{pmatrix} \\
= \; & \begin{pmatrix} c_1 I_k & c_2 I_k & \cdots & c_p I_k & c_{p+1} I_k & c_{p+2} I_k \\ 0_{k(p+1)\times k} & 0_{k(p+1)\times k} & & 0_{k(p+1)\times k} & 0_{k(p+1)\times k} & 0_{k(p+1)\times k} \end{pmatrix}.
\end{aligned}$$

Finally using (2.12), we can show that $E\left(\widehat{\Phi} - \Phi\right) = \left(c_1 + ... + c_p\right) I_k$

2.6. Tables and figures

Table 2.1. Estimated bias response surfaces b^{RS} for Models A, B, and C. Response surfaces (6) were estimated using weighted nonlinear least squares. White's heteroscedasticity-consistent standard errors are given in parentheses, $\overline{R}^2$ is the degrees-of-freedom adjusted coefficient of determination, JB is the Jarque-Bera test statistic for normality, asymptotically distributed as $\chi^2(2)$, $\star$ denotes significance at the 5% level, and $\widehat{\sigma}_u$ is the residual standard error. Coefficients and standard errors are given to 3 d.p. (to 5 d.p. for $\widehat{\beta}_6$).

	Model A	Model B	Model C
$\widehat{\beta}_1$	0.320	-3.475	-8.522
	(0.010)	(0.013)	(0.053)
$\widehat{\beta}_2$	-2.044	-1.890	-1.744
	(0.004)	(0.005)	(0.018)
$\widehat{\beta}_3$	-1.124	-1.788	-1.410
	(0.136)	(0.094)	(0.228)
$\widehat{\beta}_4$	-1.861	-1.907	-2.632
	(0.039)	(0.030)	(0.081)
$\widehat{\beta}_5$	0.999	1.038	1.404
	(0.010)	(0.009)	(0.020)
$\widehat{\beta}_6$	0.00801	0.00621	0.00240
	(0.00071)	(0.00050)	(0.00082)
$\overline{R}^2$	0.9995	0.9996	0.9976
$\widehat{\sigma}_u$	6.72	6.16	16.99
JB	1.35	$8.95^\star$	$8.92^\star$

Table 2.2. Simulated scaled bias in Models A, B, and C, for $p = 1$, and AHT and Model A approximations. All reported bias values have been multiplied by -100, b is the simulated Model A bias, b^{AHT} is the AHT approximation (3) to Model A bias, b^{RS} is the response surface approximation (6) to Model A bias, $\bar{b}$ is the simulated Model B bias, and $\widetilde{b}$ is the simulated Model C bias. In-sample points correspond to $k = 1, 2, 3, 4$ and $T = 25, 50, 100, 200$.

		VAR dimension (k)							
T		1	2	3	4	5	6	7	8
	b	6.4	13.5	20.0	26.1	31.8	37.1	42.1	46.7
	b^{AHT}	(6.4)	(12.8)	(19.3)	(25.7)	(32.1)	(38.5)	(44.9)	(51.3)
25	b^{RS}	[6.4]	[13.5]	[20.1]	[26.2]	[31.9]	[37.2]	[42.1]	[46.7]
	$\bar{b}$	19.2	25.0	30.6	35.9	40.9	45.7	50.2	54.5
	$\widetilde{b}$	35.3	40.0	44.5	49.0	53.2	57.3	61.2	64.9
	b	3.4	7.2	10.8	14.3	17.6	20.9	24.0	27.0
	b^{AHT}	(3.4)	(6.8)	(10.1)	(13.5)	(16.9)	(20.3)	(23.7)	(27.1)
50	b^{RS}	[3.3]	[7.1]	[10.8]	[14.3]	[17.8]	[21.1]	[24.3]	[27.3]
	$\bar{b}$	10.1	13.4	16.7	19.9	23.0	26.0	28.9	31.8
	$\widetilde{b}$	19.0	21.8	24.7	27.5	30.3	33.0	35.7	38.3
	b	1.7	3.7	5.6	7.5	9.3	11.1	12.9	14.6
	b^{AHT}	(1.7)	(3.5)	(5.2)	(6.9)	(8.7)	(10.4)	(12.1)	(13.9)
100	b^{RS}	[1.7]	[3.7]	[5.6]	[7.5]	[9.4]	[11.2]	[13.0]	[14.8]
	$\bar{b}$	5.2	7.0	8.7	10.5	12.2	14.0	15.7	17.3
	$\widetilde{b}$	9.9	11.4	13.0	14.6	16.3	17.9	19.5	21.1
	b	0.9	1.9	2.9	3.8	4.8	5.8	6.7	7.6
	b^{AHT}	(0.9)	(1.8)	(2.6)	(3.5)	(4.4)	(5.3)	(6.2)	(7.0)
200	b^{RS}	[0.9]	[1.9]	[2.9]	[3.8]	[4.8]	[5.8]	[6.8]	[7.7]
	$\bar{b}$	2.6	3.6	4.5	5.4	6.3	7.3	8.2	9.1
	$\widetilde{b}$	5.0	5.8	6.7	7.6	8.4	9.3	10.2	11.1
	b	0.4	0.9	1.4	1.9	2.4	2.9	3.4	3.9
	b^{AHT}	(0.4)	(0.9)	(1.3)	(1.8)	(2.2)	(2.7)	(3.1)	(3.5)
400	b^{RS}	[0.4]	[0.9]	[1.4]	[1.9]	[2.4]	[2.9]	[3.4]	[3.9]
	$\bar{b}$	1.3	1.8	2.3	2.7	3.2	3.7	4.2	4.6
	$\widetilde{b}$	2.5	3.0	3.4	3.9	4.3	4.8	5.2	5.7
	b	0.2	0.5	0.7	1.0	1.2	1.5	1.7	2.0
	b^{AHT}	(0.2)	(0.4)	(0.7)	(0.9)	(1.1)	(1.3)	(1.6)	(1.8)
800	b^{RS}	[0.2]	[0.5]	[0.7]	[1.0]	[1.2]	[1.5]	[1.7]	[2.0]
	$\bar{b}$	0.7	0.9	1.1	1.4	1.6	1.9	2.1	2.4
	$\widetilde{b}$	1.3	1.5	1.7	1.9	2.2	2.4	2.6	2.9

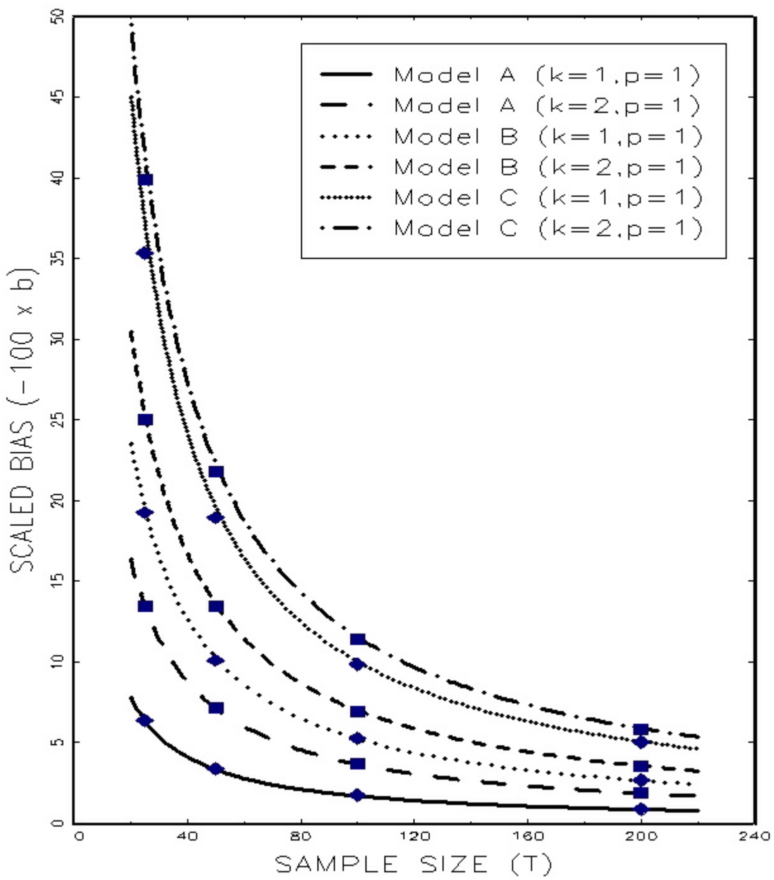
Table 2.3. Estimated variance response surfaces v^{RS} for Models A, B, and C. Response surfaces (7) were estimated using weighted non-linear least squares. White's heteroscedasticity-consistent standard errors are given in parentheses, $\overline{R}^2$ is the degrees-of-freedom adjusted coefficient of determination, JB is the Jarque-Bera test statistic for normality, asymptotically distributed as $\chi^2(2)$, $\star\star$ denotes significance at the 1% level, and $\widehat{\sigma}_u$ is the residual standard error. Coefficients and standard errors are given to 3 d.p.

	Model A	Model B	Model C
$\widehat{\gamma}_1$	-0.345	10.430	26.230
	(0.055)	(0.082)	(0.150)
$\widehat{\gamma}_2$	10.400	9.895	10.104
	(0.040)	(0.049)	(0.087)
$\widehat{\gamma}_3$	-4.469	-9.680	-17.051
	(0.203)	(0.192)	(0.250)
$\widehat{\gamma}_4$	-5.302	-4.979	-4.801
	(0.077)	(0.083)	(0.114)
$\widehat{\gamma}_5$	1.245	2.059	4.751
	(0.093)	(0.076)	(0.102)
$\widehat{\gamma}_6$	2.925	2.957	2.970
	(0.041)	(0.035)	(0.047)
$\widehat{\gamma}_7$	13.233	11.646	14.668
	(0.884)	(0.767)	(0.966)
$\widehat{\gamma}_8$	0.993	0.889	0.923
	(0.041)	(0.033)	(0.045)
$\overline{R}^2$	0.9991	0.9990	0.9982
$\widehat{\sigma}_u$	2.58	2.51	3.40
JB	$91.03^{\star\star}$	$46.38^{\star\star}$	$30.22^{\star\star}$

Table 2.4. Minimum MSE correction in Model A, for $p = 1$. λ is a correction factor, such that $\lambda\widehat{\phi}$ attains minimum MSE, br is the bias ratio≡corrected bias/OLS bias, vr is the variance ratio≡corrected variance/OLS variance $\left(\mathrm{vr} \equiv \lambda^2\right)$, bc and "$x/y$" indicate that b^2 forms x% of MSE under OLS, and *corrected* b^2 forms y% of *minimized* MSE, me is the MSE efficiency≡MSE after correction/MSE under OLS (×100). All values are computed using the appropriate response surface approximations (6) and (7). In-sample points correspond to $k = 1, 2, 3, 4$ and $T = 25, 50, 100, 200$.

		VAR dimension (k)							
T		1	2	3	4	5	6	7	8
	λ	1.05	1.12	1.19	1.26	1.32	1.39	1.46	1.52
	br	0.23	0.24	0.26	0.28	0.31	0.34	0.37	0.40
25	vr	1.11	1.25	1.41	1.58	1.75	1.94	2.13	2.32
	bc	24/1	42/3	54/5	61/7	67/10	71/13	74/15	77/19
	me	86	75	69	66	65	64	65	66
	λ	1.03	1.07	1.11	1.15	1.19	1.23	1.28	1.32
	br	0.12	0.11	0.12	0.12	0.13	0.13	0.14	0.15
50	vr	1.06	1.14	1.23	1.32	1.41	1.52	1.63	1.74
	bc	23/0.4	42/1	53/1	61/2	67/2	71/3	74/3	77/4
	me	82	67	58	52	48	45	43	42
	λ	1.02	1.04	1.06	1.08	1.10	1.12	1.14	1.16
	br	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06
100	vr	1.03	1.07	1.12	1.16	1.20	1.25	1.30	1.35
	bc	23/0.1	41/0.2	53/0.3	61/0.4	66/0.5	71/0.7	74/0.8	77/0.9
	me	80	63	53	46	41	37	34	32
	λ	1.01	1.02	1.03	1.04	1.05	1.06	1.07	1.08
	br	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03
200	vr	1.02	1.04	1.06	1.08	1.10	1.12	1.15	1.17
	bc	23/0.0	41/0.1	53/0.1	60/0.1	66/0.1	70/0.2	73/0.2	76/0.2
	me	78	61	50	43	38	34	30	28
	λ	1.00	1.01	1.01	1.02	1.02	1.03	1.04	1.04
	br	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
400	vr	1.01	1.02	1.03	1.04	1.05	1.07	1.07	1.08
	bc	23/0.0	41/0.0	52/0.0	60/0.0	66/0.0	70/0.0	73/0.0	76/0.1
	me	78	60	49	41	36	32	29	26
	λ	1.00	1.00	1.01	1.01	1.01	1.01	1.02	1.02
	br	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
800	vr	1.00	1.01	1.01	1.02	1.02	1.03	1.04	1.04
	bc	23/0.0	41/0.0	52/0.0	60/0.0	66/0.0	70/0.0	73/0.0	76/0.0
	me	77	60	48	41	35	31	28	25

Figure 2.1. Bias response surfaces: scaled bias against T, for $k = 1, 2$ and $p = 1, 2$, for Models A, B, and C. Simulated values are represented by diamonds ($k = 1$) and squares ($k = 2$)



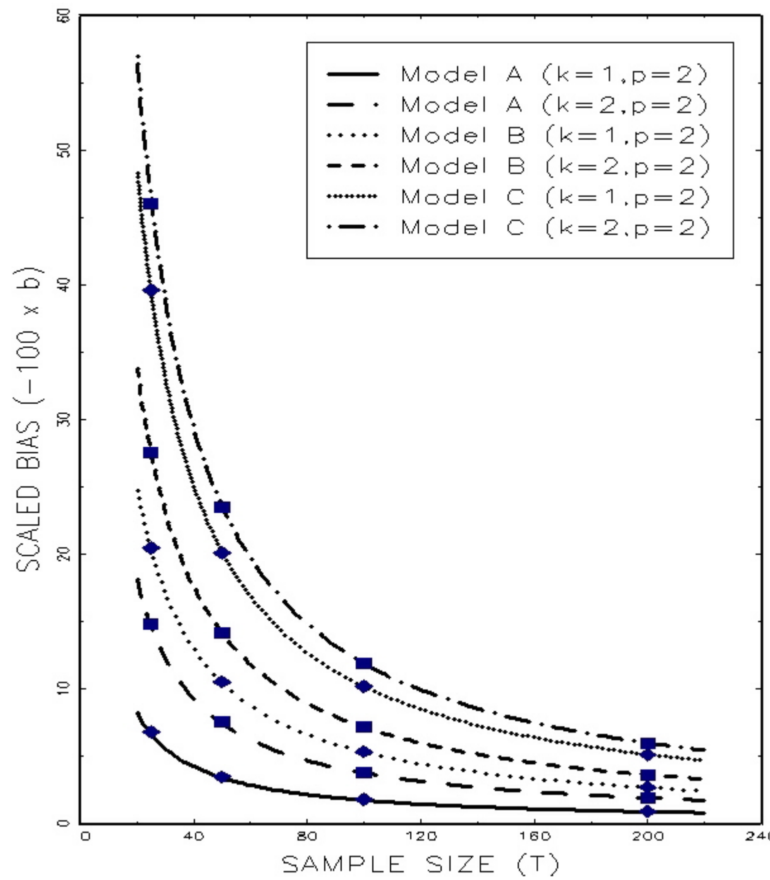
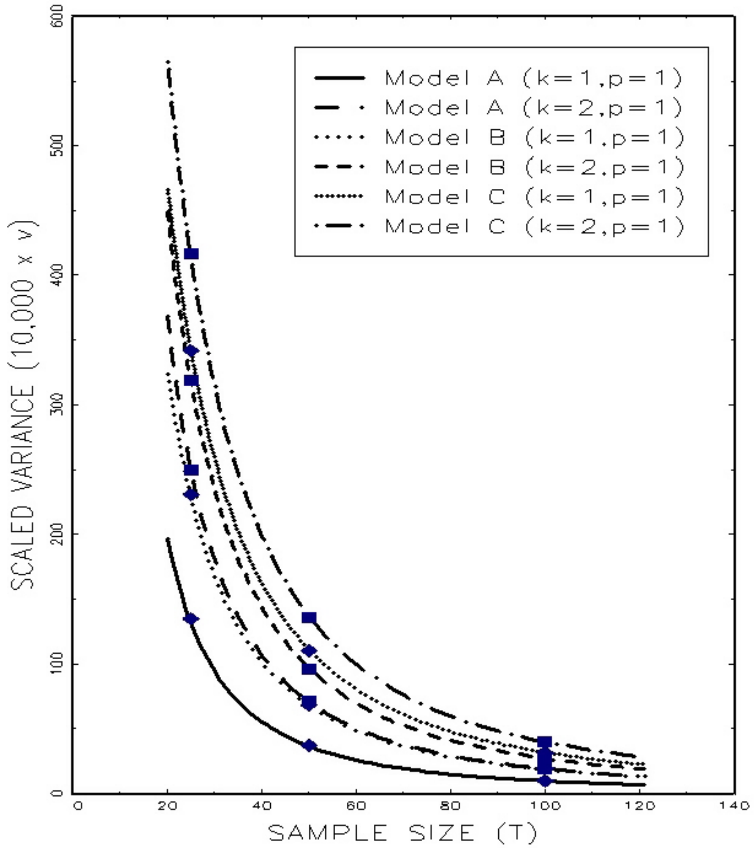
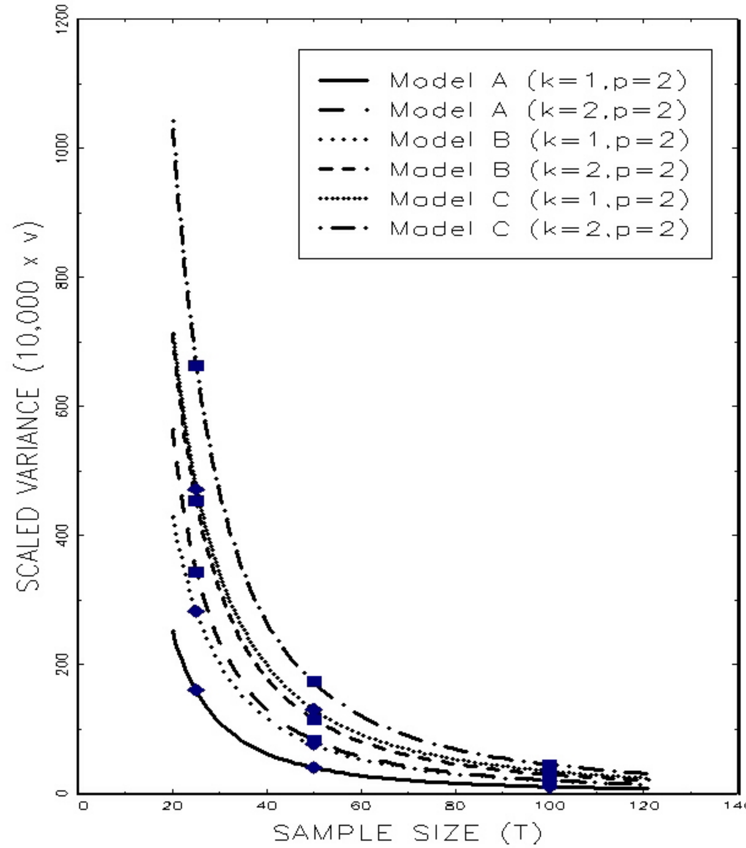


Figure 2.2. Variance response surfaces: scaled variance against T, for $k = 1, 2$ and $p = 1, 2$, for Models A, B, and C. Simulated values are represented by diamonds ($k = 1$) and squares ($k = 2$)





CHAPTER 3

Asymptotically Similar Unit Root Tests in the Presence of Autocorrelated Errors

3.1. Introduction

The unit root hypothesis has attracted a great deal of interest in econometrics. Nelson and Plosser (1982) provided empirical evidence that many macroeconomic series have a unit root. From the statistical point of view it is important to know whether or not series are stationary in order to conduct valid inference. The outcome of nonstationarity introduces the possibility of differencing the series (Plosser and Schwert, 1978) cointegration (Johansen, 1988) or error-correction (Engle and Granger, 1987) models. Banerjee *et al.* (1993) and Maddala and Kim (1998) give a review of the literature for unit root tests. Fuller (1976) and Dickey and Fuller (1979, 1981) proposed a unit root test (DF) which is widely used.

As in many testing problems, the fact that the distribution of unit root test statistics under the null hypothesis depends on nuisance parameters can result in serious size distortions for the associated unit root tests. Said and Dickey (1984) showed that the *augmented* DF (ADF) test is suitable for processes with autoregressive moving average (ARMA) errors. Phillips and Perron (1988) proposed a nonparametric testing procedure (PP) which allowed for a wider class of stationary time series in the error term. Schwert (1989) used Monte Carlo simulations to show the existence of size distortions in the ADF and PP tests. His results suggest that PP has higher power than ADF, but also much higher size distortions in the presence of negative moving average (MA) parameter in the error term. DeJong *et al.* (1992) showed that PP tests perform poorly against trend stationary alternatives and suggested the use of the Said-Dickey testing procedure.

Ng and Perron (2001) derived a class of unit root tests that take into account possible autocorrelation in the error term. The local asymptotic power function of these tests is close to the Gaussian local power envelope. They also derive the modified Akaike information criterion (MAIC) for the choice of the truncation lag. Their simulation study suggest that, for the sample sizes considered, size distortion is very low even in the presence of negative autocorrelation in the innovation sequence. These statistics are described in detail in Section 3.6. Seo (2006) pointed out that there a problem exists with the statistics derived by Ng and Perron (2001) regarding their global power: specifically, in finite samples and for alternatives far from the null, the possibility of *power reversal* occurs. Power reversal in this context means that as the true value of the parameter of interest moves farther away from the null hypothesis, power decreases. This problem is caused by the fact that the null of non-stationarity is imposed in the procedure in the construction of the modified information criteria. This type of information criteria can provide very good results with respect to control over size, but can also have serious flaws when the parameter of interest moves far from the null. Seo (2006) suggested the use of a two-step procedure in which he first fits an autoregression to get the estimated residuals and at the second step uses them as a proxy for the MA component. Perron and Qu (2007) address the issue of power reversal and improve the performance of the statistics by introducing a two step procedure, using OLS estimation for the choice of the order of the lagged differenced terms and GLS estimation for the calculation of the statistics. As can be seen from their results (Figures 1-4), the problem becomes less severe, but is still evident for the case of no autocorrelation in the error term.

This Chapter addresses the issue of unit root testing in the presence of correlated innovation errors that take the form of a finite order moving average process. Following Hillier (1987), our approach is based on obtaining a characterization of the class of *similar tests*. These are tests whose size does not depend on nuisance parameters, provided that sufficient statistics for the nuisance parameters exist, under the null hypothesis. Given the fact that a sufficient statistic for the MA parameters is not available, we consistently estimate the MA parameters by maximum likelihood, and then use the above estimates to characterize the class of (asymptotically) similar tests. After the characterization of the class of similar regions we proceed to the selection of some tests within this class by the use of appropriate optimality criteria. The advantage of such an approach is that we can focus our attention on a set of tests whose exact size is independent of the nuisance parameters involved. In this way we can address the serious issue of size stability at the first stage of selecting a test.

In order to choose statistics from the class of asymptotic similar tests we make use of the optimality criteria proposed by Forchini and Marsh (2000). They derive unit root tests according to the Bounded Norm Minimizing (BNM) and Bounded Estimated Point Optimal (BEPO) criteria under the assumption of i.i.d. innovation errors. We apply the same optimality criteria to derive tests statistics in a more general framework that allows the presence of possibly correlated innovation errors that may take the form of a finite order MA process. The objective is to derive unit root tests with fairly stable size over MA processes with varying order and values of associated parameters, and with high global power in comparison to other unit root tests existing in the literature.

The Chapter is organized as follows. In Section 3.2 we refer to the theory related to the construction of similar tests. Section 3.3 describes the BNM and BEPO optimality criteria for the choice test statistics proposed by Forchini and Marsh (2000). Section 3.4 describes the construction of similar regions in the case of correlated errors and in Section 3.5 we use the optimality criteria to derive the test statistics followed by the description of the method of estimation we are using. The limiting distributions of the resulting test statistics are derived in the presence of deterministics consisting of an intercept term only and an intercept and a linear trend. In Section 3.6 the finite-sample performance of the statistics is assessed in the context of a simulation study. In Section 3.7 we provide some concluding remarks. All proofs are included in the technical Appendix of Section 3.8. Tables and figures are presented in the last Section of the Chapter.

3.2. Methodology on the characterization of similar regions

The methodology we follow for the characterization of similar regions is described by Hillier (1987). Let z be a vector of random variables with density $f(z;\eta,\theta)$ depending on two vectors of parameters η, and θ. If we want to test the null hypothesis

$$H_0 : \theta = \theta_0$$

then θ is the vector of parameters of interest and η is the vector of nuisance parameters. In general the size of any critical region ω in this context will be dependent on η; i.e.,

$$\int_\omega f(z;\eta,\theta_0) = \alpha(\eta).$$

Critical regions related to this problem which are independent of nuisance parameters

$$\int_\omega f(z;\eta,\theta_0) = \alpha$$

are called *similar* critical regions. If there is a sufficient statistic t for η under H_0 the density function is given by

$$f(z;\eta,\theta_0) = pdf(t;\eta,\theta_0)pdf(z|t;\theta_0)$$

where $pdf(t;\eta,\theta_0)$ is the density of the sufficient statistic under H_0 and $pdf(z|t;\theta_0)$ is the conditional density of z given t, which is independent of the nuisance parameter η. So, provided we have sufficient statistics for η, the conditional distribution of z given these statistics will be free of nuisance parameters and will result to a similar critical region.

Hillier (1987) summarizes the procedure for constructing similar regions in the following theorem.

Let t be a boundedly complete sufficient statistic for the nuisance parameter η under H_0. If for almost all t there is a one-to-one transformation $z \longmapsto (t(z), v(z))$ for which under H_0 v is independent of t, then the statistic v characterizes the class of similar regions for testing H_0 in the sense that a region ω is similar of size α if and only if ω has size α in the distribution of v.

We are going to use this theorem for the characterization of similar critical tests in the case where the innovation sequence takes the form of an MA process.

3.3. Optimality criteria

We now address the question of how to select a particular test from within the class of similar tests. Ideally, we would choose a *Uniformly Most Powerful* (UMP) test. A UMP test is a test which has the highest available power for every η, and θ. In unit root tests the power of a test depends on the nuisance parameters η and the value of the parameter of interest θ under H_1, so it is not possible to achieve the UMP criterion. Consequently, we have to use weaker optimality criteria for the selection of a test. Cox and Hinkley (1974) suggest some alternative optimality criteria, such as the selection of a typical alternative for θ (*point optimal* (PO)) or the construction of a *locally most powerful* (LMP) test, which involves the maximization of the power of the test in the neighborhood of the null hypothesis. Selecting a typical value of θ could be seen as arbitrary unless there is specific prior information for the parameter. The problem with the LMP tests is that their power can often be low, especially for alternatives that lie far from the null (Zaman, 1996, pp. 133-136).

Forchini and Marsh (2000) suggest the use of two alternative optimality criteria. Their statistical framework can be summarized as follows. Consider a $N \times 1$ vector of observables and a vector of unknown parameters $(\theta, \sigma^2) \in \mathbb{R} \times (0, \infty)$. The null hypothesis $H_0 : y \sim N(0, \sigma^2\Omega(\theta_0))$ is tested against $H_1 : y \sim N(0, \sigma^2\Omega(\theta))$ using the critical region resulting from the rule reject H_0 if

$$\frac{y'\Omega^{-1}(\theta)y}{y'\Omega^{-1}(\theta_0)y} < k_\alpha \tag{3.1}$$

where k_α is chosen so that α is the size of the test. It is clear that when the numerator changes with θ there is no a UMP test.

In the absence of a UMP test two weaker optimality criteria are presented in Sections 3.3.1 and 3.3.2 below.

3.3.1. Bounded Norm Minimizing tests

Suppose that $y'\Omega^{-1}(\theta)y \leq l(\theta)'\Psi(y)l(\theta)$, where $l(\theta)$ is a vector depending only upon θ and $\Psi(y)$ is a positive definite matrix depending only upon y.

A sufficient condition for

$$\frac{l(\theta)'\Psi(y)l(\theta)}{y'\Omega^{-1}(\theta_0)y} < k_\alpha$$

is to minimize the norm

$$\left\| \frac{\Psi(y)}{y'\Omega^{-1}(\theta_0)y} \right\| < k,$$

for k such that the size of the test is α. The norm in the above equation can be any matrix norm (see e.g. Horn and Johnson, 1985). Notice that any norm of the matrix $\Psi(y)/y'\Omega^{-1}(\theta_0)y$ gives a *norm minimizing* (NM) test and when (3.1) holds with equality and a BNM test when the inequality is strict.

3.3.2. Bounded Estimated Point Optimal Tests

The second optimality criterion is that of using *estimated point optimal* tests (EPO). This criterion is related to the PO tests which are discussed above. Even if the alternative is generally unknown, it is possible to estimate it with the value θ^* which satisfies

$$\theta^* = \arg\min_\theta \left\{ \frac{l(\theta)'\Psi(y)l(\theta)}{y'\Omega^{-1}(\theta_0)y} \right\}$$

for a set of observations y. In the case where (3.1) holds with equality, the EPO critical region is given by

$$\frac{l(\theta^*)'\Psi(y)l(\theta^*)}{y'\Omega^{-1}(\theta_0)y} < k, \tag{3.2}$$

where k is chosen such that the size of the test is α. As with the case of the BNM criterion, if (3.1) does not hold with equality, (3.2) is a BEPO test. Another criterion of this type is to reject H_0 if

$$|\theta^* - \theta_0| > k_a, \tag{3.3}$$

where θ_0 is the value of the parameter under H_0 and k_a is chosen such that the size of the test is a.

Forchini and Marsh (2000) use the above criteria for the derivation of similar unit root test statistics. Simulation results suggest that these statistics have distorted size in the presence of an MA(1) error. In the presence of an MA process in the errors, these test statistics are no longer similar due to the fact that their critical regions depend on the associated MA parameters. The approach in this Chapter is to modify the construction of the UMP critical region in order to take into account the possibility of an MA(m) process in the errors. Then we apply the BNM and BEPO optimality criteria to choose statistics from the class of asymptotically similar tests and we find that these have good power properties in finite samples.

3.4. Construction of similar critical regions

Marsh (2005) considers a linear regression model with an MA term in the errors and characterizes the class of asymptotically similar tests. We use the BNM and

BEPO optimality criteria for deriving tests within this class. The model is

$$y = X\beta + u, \tag{3.4}$$

where β is a $k \times 1$ vector of parameters, X a $N \times k$ full rank matrix of the deterministic components (in this Chapter we consider an intercept and a linear trend), $u = (u_1, ..., u_N)'$ and

$$\begin{aligned} u_t &= \rho u_{t-1} + \zeta_t \\ \zeta_t &= \sum_{j=0}^{m} \phi_j \varepsilon_{t-j} \\ \varepsilon_t &\sim NIID(0, \sigma^2) \end{aligned}$$

for $t = 1, ..., N$, $u_0 = 0$, and $\phi_0 = 1$. We impose the invertibility condition $\left|\phi_j\right| < 1$ for $j = 1, ..., m$. So the parameters involved are $\theta = (\rho, \beta', \sigma^2, \phi')$ with parameter space $\Theta = (-1, 1] \times \mathbb{R}^k \times \mathbb{R}^+ \times (-1, 1)^m$.

In the context of (3.4) the unit root hypothesis takes the form

$$H_0 : \rho = 1 \text{ vs. } H_1 : |\rho| < 1,$$

with β, σ^2 and ϕ the nuisance parameters for this testing problem. The method described in Section (3.2) is going to be applied for the construction of similar critical region for the hypothesis stated above. Invariant transformations are applied on the data y, which do not affect the decision with respect to H_0 and H_1, but take out the effect of the nuisance parameters. These transformations involve the use of some matrices defined below.

Let $L^{(i)}$ be the lower-triangular matrix with ones on the i^{th} off-diagonal and zeros elsewhere. Multiplying (from any side) $L^{(i)}$ by any vector gives the i^{th} lag of this vector leaving the first element of the vector unchanged. For this reason we refer to $L^{(i)}$ as the lag-matrix. Using $L^{(i)}$, T_ρ is defined as

$$T_\rho = (I_N - \rho L^{(1)}). \tag{3.5}$$

Note therefore that $T_1 = I_N - L^{(1)}$. Multiplying any vector by T_1 results the vector of first differences for the last $N-1$ elements leaving the first element unchanged (implicitly a zero initial condition is imposed). So T_1 acts as a first difference operator that transforms an $I(1)$ series to $I(0)$ except from the the first element which remains unchanged and is asymptotically negligible.

Then, using the $L^{(i)}$ matrix again K_ϕ is defined as

$$K_\phi = (I_N + \sum_{i=1}^{m} \phi_i L^{(i)}). \tag{3.6}$$

So when the K_ϕ matrix is multiplied by a vector of white noise errors this results in an MA vector series of order m. Using this rationale K_ϕ^{-1} transforms a vector of MA(m) to a vector of white noise series. Defining

$$\phi = (\phi_1, ..., \phi_m)' \text{ and } \varepsilon = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_N)',$$

(3.4) can be expressed as

$$T_\rho(y - X\beta) = K_\phi\varepsilon. \tag{3.7}$$

At this point, the transformation matrices listed above are used to clear the distribution of the vector of observables from the nuisance parameters. We start from the joint sample density of y, which is

$$y \sim N(X\beta, \sigma^2 T_\rho K_\phi K_\phi' \left(T_\rho^{-1}\right)').$$

Then, for notational simplicity, we define

$$x = K_\phi^{-1} T_1 y, \tag{3.8}$$

$$Z = K_\phi^{-1} T_1 X, \tag{3.9}$$

and (3.7) is transformed under H_0 to

$$x = Z\beta + \varepsilon.$$

The distribution of x is given by

$$x \sim N(Z\beta, \sigma^2 \Sigma_{\rho,\phi})$$

where

$$\Sigma_{\rho,\phi} = K_\phi^{-1} T_1 T_\rho^{-1} K_\phi K_\phi' \left(T_\rho^{-1}\right)' T_1' \left(K_\phi^{-1}\right)'. \tag{3.10}$$

Note that x under H_0 is

$$x \sim N(Z\beta, \sigma^2 I_N)$$

At this point it is useful to use the following lemma before proceeding.

Lemma 3.1. *The matrix $\Sigma_{\rho,\phi}$ given in (3.10) can be expressed as*

$$\Sigma_{\rho,\phi} \equiv \Sigma_\rho = T_1 T_\rho^{-1} \left(T_\rho^{-1}\right)' T_1'.$$

For the characterization of the class of similar tests the methodology by Hillier (1987) described in Section 2 is applied in this setup. Using the Cholesky decomposition, the projection matrix

$$M_Z = I_N - Z\left(Z'Z\right)^{-1} Z'$$

can be decomposed as:

$$\begin{aligned} CC' &= M_Z \\ C'C &= I_{N-k} \end{aligned}$$

where C is a $N \times N - k$ matrix.

The following transformation are applied using C matrix. First x is transformed as

$$x \longmapsto \begin{pmatrix} \widehat{\beta} = \left(Z'Z\right)^{-1} Z'x \\ w = C'x \end{pmatrix}$$

and then w as

$$w \longmapsto \left(\begin{array}{l} s^2 = \|w\|^2 = x' M_Z x \\ \upsilon = \frac{w}{\|w\|} = C' x / s \end{array} \right)$$

As it can be seen from the above, $\widehat{\beta}$ is not feasible due to the fact that is dependent on ϕ. It is possible however to proceed by finding a consistent estimate of ϕ.

The distribution of w is

$$w \sim N(0, \sigma^2 C' \Sigma_\rho C) \overset{H_0}{\sim} N(0, \sigma^2 I_{N-k}) \tag{3.11}$$

Marsh (2007) gives the density of υ with respect to the normalized Haar measure on the surface of the unit $N-k$ sphere to be

$$pdf(\upsilon) = \det \left(C' \Sigma_\rho C \right)^{-1/2} \left[\upsilon' \left(C' \Sigma_\rho C \right)^{-1} \upsilon \right]^{-\frac{N-k}{2}}, \tag{3.12}$$

According to the above, the most powerful critical region of H_0 vs. H_1 has critical region given by

$$\upsilon' \left(C' \Sigma_\rho C \right)^{-1} \upsilon < k_\alpha, \tag{3.13}$$

where k_α is chosen such that the size of the test is α. Note that, in view of Lemma 3.1, the above critical region is independent of the MA parameters of the vector ϕ under both the null and the alternative hypothesis.

3.5. Asymptotically similar statistics

After the characterization of the class of asymptotically similar statistics we use the optimality criteria suggested by Forchini and Marsh (2000) in order to derive test statistics from this class. Since there is not a sufficient statistic for the MA parameters included in matrix K_ϕ, these parameters are estimated using maximum likelihood estimation (MLE). The matrix K_ϕ including the estimated MA parameters is denoted as $K_{\widehat{\phi}}$. More explicitly we define

$$Z_{\widehat{\phi}} = K_{\widehat{\phi}}^{-1} T_1 X \text{ and } \nu = M_{Z_{\widehat{\phi}}} K_{\widehat{\phi}}^{-1} T_1 y. \tag{3.14}$$

The procedure that gives the order of the MA process and the estimation of the MA parameters is described in detail later in Section 3.5.1. We define

$$\Psi_{11} = \left(T_1^{-1}\right)' T_1^{-1} \tag{3.15}$$

$$\Psi_{12} = \left(T_1^{-1}\right)' \left(T_1^{-1} - I_N\right) \tag{3.16}$$

$$\Psi_{22} = \left(T_1^{-1} - I_N\right)' \left(T_1^{-1} - I_N\right). \tag{3.17}$$

Theorem 3.1. *Let $\|.\|$ denote a norm on the space 2×2 positive definite matrices, and let*

$$\Psi(\nu) = \frac{1}{\nu'\nu} \begin{pmatrix} \nu'\Psi_{11}\nu & \nu'\Psi_{12}\nu \\ \nu'\Psi_{12}\nu & \nu'\Psi_{22}\nu \end{pmatrix} \tag{3.18}$$

Then a BNM test is: reject $H_0 : \rho = 1$ *if*

$$N^{-1} \|\Psi(\nu)\| < k_\alpha \tag{3.19}$$

where ν *is defined in (3.14) and* k_α *is chosen such that the size of the test is* α*.*

Theorem 3.1 generates a class of BNM tests, depending upon the choice of particular norm. A statistic from this class could result from the use of the Euclidean matrix norm $\|\Psi(\nu)\| = \{tr\Psi(\nu)'\Psi(\nu)\}^{1/2}$ or the spectral norm of $\Psi(\nu)$, defined as the square root of the maximal eigenvalue of $\Psi(\nu)'\Psi(\nu)$.

Theorem 3.2. *A BEPO test for* $H_0 : \rho = 1$ *against* $H_1 : -1 < \rho < 1$ *is given by the following rule:*

reject H_0 *if*

$$BEPO = N \left| \frac{\nu'\Psi_{12}\nu - \nu'\Psi_{22}\nu}{\nu'\Psi_{22}\nu} \right| > k_\alpha \tag{3.20}$$

where ν *is defined in (3.14) and* k_a *is such that the size of the tests is* α*.*

3.5.1. Estimation of the MA process

Both the BNM and BEPO statistics contain the matrix $K_{\widehat{\phi}}$ of estimated MA coefficients. The construction of this matrix requires two steps: a procedure that detects the order of the MA component and an estimation method for the MA parameters. Treating both these aspects as a priori unknown makes the inference of Theorems 3.1 and 3.2 asymptotically feasible and suitable for practical application.

We first discuss the estimation of the MA parameters for a given order. In the absence of a sufficient statistic for ϕ, we need to employ a consistent estimator. It has to be stressed that the choice of a good estimator for ϕ is of major importance for the good properties (empirical size near to the nominal one and high power) of the statistics. We estimate ϕ by conditional maximum likelihood or pseudo-maximum likelihood if we do not wish to maintain the normality assumption on the innovation errors. It is a well known fact that, under the invertibility assumption imposed on the moving average process, the (pseudo) maximum likelihood estimator of ϕ is $\sqrt{N}$-consistent. (see Anderson, 1971).

Having estimated models of certain order m, we use information based rules to choose one among them. These are the criteria proposed by Akaike (1974), Schwarz (1978) and Hannan and Quinn (1979), denoted henceforth as AIC, BIC and HQIC respectively. These are described in detail below.

The algorithm for estimating ϕ is described below. We first estimate the following model with least squares:

$$y_t = X\widehat{\beta} + \widehat{u}_t, \tag{3.21}$$

where the deterministic component X includes an intercept only, or an intercept and a trend. We then fit the following $ARMA(1,m)$ model to the residuals of (3.21)

$$\widehat{u}_t = \rho\widehat{u}_{t-1} + \varepsilon_t + \sum_{i=1}^{m} \phi_i \varepsilon_{t-i},$$

for $t = 1, 2, ..., N$. We set a minimum value $m_{\min}$, and a maximum value $m_{\max}$ for the order of the MA component. We estimate $ARMA(1, m)$ models with $m_{\min} \leq m \leq m_{\max}$. For each model, we condition on the m first values of ε being zero:

$$\varepsilon_0 = \varepsilon_1 = ... = \varepsilon_m = 0.$$

From the above assumptions we can iterate on:

$$\varepsilon_t = (\widehat{u}_t - \rho \widehat{u}_{t-1}) - \sum_{i=1}^{m} \phi_i \varepsilon_{t-i},$$

for $t = 1, 2, ..., N$.

The conditional log likelihood is

$$\mathcal{L}\left(\rho, \phi, \sigma^2\right) = -\frac{N}{2} \log(2\pi) - \frac{N}{2} \log(\sigma^2) - \sum_{t=1}^{N} \frac{\varepsilon_t^2}{2\sigma^2}.$$

Since we assumed $\left|\phi_j\right| < 1$ for $j = 1, ..., m$ the effect of the initial condition fades out as sample size increases (Hamilton, 1994, p.128).

After the estimation of $m_{\max} - m_{\min} + 1$ models we use information criteria to choose one of them. These information criteria are the following

$$\begin{aligned} IC_{AIC}(m) &= -2\frac{\mathcal{L}}{N} + \frac{2(m+1)}{N} \\ IC_{BIC}(m) &= -2\frac{\mathcal{L}}{N} + \frac{(m+1)\ln(N)}{N} \\ IC_{HQIC}(m) &= -2\frac{\mathcal{L}}{N} + \frac{2(m+1)\ln(\ln(N))}{N}. \end{aligned}$$

We choose m such that the information criterion (used in each case) is minimized:

$$\widehat{m} = \arg\min_{m} IC(m).$$

After choosing the order of the MA component and estimating the MA parameters, we can substitute them in the sufficient statistics for (β, σ^2) and then construct the similar critical regions. It is important to note here that, asymptotically, the test statistics we derive do not depend on the nuisance parameter under H_0 since $Z = K_{\hat{\phi}}^{-1} T_1 X$ and $T_1 u = K_{\phi}\varepsilon$ which gives

$$\begin{aligned} \nu &= M_Z K_{\hat{\phi}}^{-1} T_1 \left(X\beta + u \right) = M_Z K_{\hat{\phi}}^{-1} T_1 u \\ &= M_Z K_{\hat{\phi}}^{-1} K_{\phi}\varepsilon = \left[I + o_p\left(1\right) \right] M_Z \varepsilon. \end{aligned}$$

The above result shows that the statistics we derive are asymptotically similar.

3.5.2. Limiting distribution of BNM and BEPO statistics

Having derived the BNM and BEPO test statistics for the unit root hypothesis, we proceed to derive their limiting distributions. To this end, we restrict the deterministic components of the data generating process to an intercept and a linear trend, i.e. we assume that the matrix of deterministics in (3.4) takes the form

$$X' = \begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & 2 & \dots & N \end{pmatrix}, \tag{3.22}$$

or

$$X' = \left(\begin{array}{cccc} 1 & 1 & \ldots & 1 \end{array} \right), \tag{3.23}$$

which corresponds to the case where only an intercept is included in the model.

Theorem 3.3. *Consider the process in (3.4) and let $W(\cdot)$ be standard Brownian motion on $D[0,1]$. Under the null hypothesis $H_0 : \rho = 1$ the following limit theory applies as $N \to \infty$:*

for X satisfying (3.22):

(i) *The BNM test of Theorem 3.1 satisfies*

$$BNM \Rightarrow 2\left\{ \int_0^1 W^2(r)dr - 2W(1)\int_0^1 rW(r)dr + \frac{1}{3}W^2(1) \right\}.$$

(ii) *The BEPO test of Theorem 3.2 satisfies*

$$BEPO \Rightarrow \frac{1}{2}\frac{1}{\left| \int_0^1 W^2(r)dr - 2W(1)\int_0^1 rW(r)dr + \frac{1}{3}W^2(1) \right|}.$$

For X satisfying (3.23):

(iii) *The BNM test of Theorem 3.1 satisfies*

$$BNM \Rightarrow 2\left\{ \int_0^1 W^2(r)dr \right\}.$$

(iv) *The BEPO test of Theorem 3.2 satisfies*

$$BEPO \Rightarrow \frac{1}{2}\left| \frac{W^2(1) - 1}{\int_0^1 W^2(r)dr} \right|.$$

3.6. Numerical Study

The test statistics we develop are motivated asymptotically in the sense that they are asymptotically similar with respect to the MA parameter. In order to examine their size and power properties in small samples we employ a Monte Carlo study. Two models are considered for the simulations: the first is based on (3.4) with X defined as in (3.22) for the case of a constant and trend included and (3.23) for the case of a constant only included. The DGP used for the simulations has the following specification:

$$\begin{aligned} u_t &= \rho u_{t-1} + \varepsilon_t + \phi \varepsilon_{t-1}, \\ \varepsilon_t &\sim NIID(0,1). \end{aligned}$$

Each Monte Carlo experiment is based on 10000 replications. We investigate the size distortion and power of the statistics in finite samples. For the numerical study related to the size distortion, the following minimal complete factorial design is used with values for the parameters:

$$\begin{aligned} \phi &= -0.8, -0.7, ..., 0.8, \\ N &= 50, 100, 200, 400, \\ \rho &= 1, \\ \alpha &= 0.05, \end{aligned}$$

where α is the nominal size of the test statistics.

For the numerical study investigating the finite sample power of the statistics, the simulation design includes all combinations of the following parameter values:

$$\begin{aligned} \rho &= 0.8, 0.82, ..., 0.98, \\ N &= 50, 100, 200, 400, \\ \phi &= -0.5, 0, \\ \alpha &= 0.05, \end{aligned}$$

and

$$\begin{aligned} \rho &= 0.1, 0.2, ..., 0.7, \\ N &= 50, 100, 200, 400, \\ \phi &= 0, \\ \alpha &= 0.05. \end{aligned}$$

The statistics BNM_0 and $BEPO_0$ correspond to the case in which MA terms are not estimated. These are the statistics proposed by Forchini and Marsh (2000). In order to construct statistics BNM_0 and $BEPO_0$ we set $\phi = 0$ (i.e. $K_\phi = I_N$) in (3.19) and (3.20). For the BNM_a and $BEPO_a$ statistics the AIC is used, for the BNM_b and $BEPO_b$ the BIC is used, and for BNM_h and $BEPO_h$ the HQIC is employed. We refer to these test statistics as similar statistics. The information criteria consider MA(m) processes with $m_{\min} = 0$ and $m_{\max} = 5$. Initially, we use exact critical values for the BNM and BEPO statistics. Later in the discussion we examine the behaviour of BNM and BEPO statistics that use asymptotic critical values.

We compare the finite sample performance of the statistics derived in this Chapter with other statistics in the literature. In Ng and Perron (2001) the following statistics can be found:

$$
\begin{aligned}
MZ_a^{GLS} &= \frac{N^{-1}\tilde{y}_N^2 - s_{AR}^2}{2N^{-2}\sum_{t=1}^{N}\tilde{y}_{t-1}^2}, \\
MSB^{GLS} &= \left(\frac{N^{-2}\sum_{t=1}^{N}\tilde{y}_{t-1}^2}{s_{AR}^2}\right)^{\frac{1}{2}} \\
MZ_t^{GLS} &= MZ_a^{GLS} \times MSB^{GLS},
\end{aligned}
$$

where $\tilde{y}_t = y_t - x_t\widehat{\gamma}^{GLS}$, ($x_t$ being the t-th row of X) and $\widehat{\gamma}^{GLS}$ being the GLS estimate of γ. This is calculated by the GLS regression of $y_t^{\overline{a}}$ on $x_t^{\overline{a}}$, where $y_t^{\overline{a}} = y_t - \overline{a}y_{t-1}^{\overline{a}}$ for $t = 2, ..., N$ and $y_1^{\overline{a}} = y_1$. Following Elliot *et al.* (1996), for the case of a constant only in the model $\overline{a} = 1 - \frac{7}{N}$, and when a constant and trend are included $\overline{a} = 1 - \frac{13.5}{N}$.

Ng and Perron (2001) also modify the feasible point optimal test suggested by Elliot *et al.* (1996) which is:

$$P_T = \frac{S\left(\overline{a}\right) - \overline{a}S\left(1\right)}{s_{AR}^2},$$

where $S(a) = \inf_\gamma \sum_{t=1}^{N} \left(y_t^a - \gamma x_t^a\right)^2.$

The modified feasible point optimal test suggested by Ng and Perron (2001) for the constant case is:

$$MP_T^{GLS} = \frac{\overline{c}^2 N^{-2} \sum_{t=1}^{N} \tilde{y}_{t-1}^2 - \overline{c} N^{-1} \tilde{y}_N^2}{s_{AR}^2},$$

and for the case of a constant and trend included in the deterministics is:

$$MP_T^{GLS} = \frac{\overline{c}^2 N^{-2} \sum_{t=1}^{N} \tilde{y}_{t-1}^2 + \left(1 - \overline{c}\right) N^{-1} \tilde{y}_N^2}{s_{AR}^2}.$$

The autoregressive spectral density estimate of σ^2 is defined as:

$$\begin{aligned} s_{AR}^2 &= \frac{\hat{\sigma}_{ek}^2}{\left(1 - \sum_{t=1}^{N} \hat{b}_i\right)^2}, \\ \hat{\sigma}_{ek}^2 &= N^{-1} \sum_{t=k+1}^{N} \hat{e}_{tk}^2, \end{aligned}$$

with $\hat{b}_i$ and e_{tk}^2 derived from the following OLS regression:

$$\Delta\tilde{y}_t = \hat{b}_0\tilde{y}_{t-1} + \sum_{i=1}^{k}\hat{b}_i\Delta\tilde{y}_{t-i} + \hat{e}_{tk}.$$

Note that the above regression is used for the GLS ADF test. More specifically a t-test is run on $H_0 : b_0 = 0$.

The Modified Akaike Information Criterion used for the determination of the autoregressive order k is:

$$MAIC(k) = \ln\left(\hat{\sigma}_k^2\right) + 2\frac{\tau_T\left(k\right) + k}{N - k_{\max}},$$

where

$$\begin{aligned}
\tau_T\left(k\right) &= \frac{\hat{b}_0 \sum_{t=k_{\max}+1}^{N} \tilde{y}_{t-1}^2}{\hat{\sigma}_k^2}, \\
\hat{\sigma}_k^2 &= \frac{\sum_{t=k_{\max}+1}^{N} \hat{e}_{tk}}{N - k_{\max}}.
\end{aligned}$$

The upper bound is set to $k_{\max} = int\left(12(N/100)^{1/4}\right)$. The value of k chosen by $MAIC(k)$ is such that $k = \arg\min_{k\in[0,k_{\max}]}$.

In the tables of this Chapter, MZ_a and MZ_t are the modified PP statistics and MSB is the modified Sargan-Bhargava statistic. P_T refers to the feasible point optimal test and MP_T to its modified variant. All these statistics use GLS detrending. ADF corresponds to the ADF statistic with GLS detrending and for

ADF_{LS}, OLS detrending is used. For both ADF and ADF_{LS} the MAIC is used. MZ_{aLS} statistic denotes the MZ_a statistic based on OLS detrending. Lastly, MZ_{a2} corresponds to the MZ_a statistic with GLS detrending used for the data and OLS detrending used for the spectral density estimation.

Tables 3.2 ($N = 50, 100$) and 3.3 ($N = 200, 400$) report size distortion of the statistics for a model including an intercept term only (X defined by (3.23)) and Tables 3.4 ($N = 50, 100$) and 3.5 ($N = 200, 400$) report the size distortion of the statistics for the case of an intercept and a trend included in the model (X defined by (3.22)). A first observation is that serious size distortions occur when the MA parameter is specified to be near to -1. Tables 3.2-3.5 show that the statistics derived in this Chapter exhibit much lower size distortion in comparison to the BNM_0 and $BEPO_0$ statistics. It can be also seen that the choice of the specific information criterion is crucial for the level of size distortion in small samples ($N = 50, 100$). More specifically, the size distortion for the BNM and BEPO statistics is the lowest when the AIC is used. When the HQIC is used, size distortion becomes higher and the use of BIC gives the highest size distortion among all information criteria considered for our statistics.

The relatively good performance of BNM_a and $BEPO_a$ with respect to size distortion could be explained by the fact that the AIC is the most *liberal* (tends to choose comparatively higher order for the MA process) of all information criteria. This is evident in Figure 3.1 which presents the relative frequencies of the order chosen by each information criterion under H_0 for different values of ϕ (for a model with an intercept and a trend, sample size $N = 100$). It can be seen that BIC is the most *conservative* information criterion, in the sense that, keeping everything else constant, it tends to choose the lowest MA order in comparison to the other two criteria. This has a very detrimental effect (with respect to control over size) for values of ϕ close to -1 and that is why BNM_b and $BEPO_b$ give the highest size distortion in small samples. It is also observed that for sample sizes $N = 200, 400$, the different information criteria deliver almost the same empirical size.

Comparing the statistics derived in this Chapter with other statistics in the literature, we find that they have much lower size distortion for $N = 50$. For values of the MA parameter being close to -1, it is obvious that all the the statistics of Ng and Perron (2001) have extremely high size distortion, making them unreliable for such small sample sizes. This is important as sample sizes of this kind are relevant in applied research. For higher sample sizes, statistics MZ_a, MZ_t, MSB, P_T and MP_T appear to have very small size distortion and perform better than the similar statistics. Figures 3.2 and 3.3 illustrate graphically the facts mentioned above.

Another crucial observation for the similar statistics is that their size distortion reduces as the sample size N increases. For example, in the case of a model with an intercept only (Tables 3.2 and 3.3), when $\phi = -0.8$, $BEPO_a$ statistic has size 0.367 for $N = 50$, 0.22 for $N = 100$, 0.101 for $N = 200$, and 0.07 for $N = 400$. We observe the same behaviour for BNM_b, $BEPO_b$, BNM_h and $BEPO_h$. This observation suggests that the empirical size of the similar statistics derived in this

Chapter converges to its nominal value (5% in this case), as sample size increases. This can be attributed to the consistency of the maximum likelihood estimator as well as the better performance of the information criteria as N increases. This is not the case for statistics BNM_0 and $BEPO_0$: size distortion increases as sample size N increases. The $BEPO_0$ statistic for example has size 0.673 for $N = 50$, 0.832 for $N = 100$ and 0.908 for $N = 200$, when $\phi = -0.8$. This suggests that empirical size of the BNM_0 and $BEPO_0$ statistics can go farther from nominal size as N increases in the presence of autocorrelation in the errors. In the case of an intercept and a trend included in the model (Tables 3.4 and 3.5) we observe that the level of size distortion increases for all the statistics.

We also observe that the size of the Ng and Perron statistics that use the MAIC does not always move closer to the nominal size as sample size increases. For example, Tables 3.2 and 3.3 show that the empirical size for $\phi = -0.8$ of MZ_t is 0.988 for $N = 50$, 0.039 for $N = 100$, 0.021 for $N = 200$ and 0.027 for $N = 400$. At first view this could be considered as not not necessarily problematic, since one would be interested to get size less or equal to the nominal one. However, in cases that empirical size appears to be substantially lower than the nominal one there could be detrimental effects on the power of the statistic. This problem could be "hidden" in cases that size-adjusted power results are presented. In cases in which the empirical size is substantially lower than the nominal one, we would expect size-unadjusted power to be lower than size-adjusted power, at least for alternatives close to the null.

Tables 3.6 and 3.7 report the power of the statistics for models corresponding to X defined by (3.23) and (3.22) respectively, when there is no autocorrelation in the error term ε_t ($\phi = 0$). This is not a favourable case for the statistics we derive in this Chapter, since MA processes are considered which do not exist under the data generating process. However, we observe that the power of BNM_b and $BEPO_b$ statistics is very close to the power of BNM_0 and $BEPO_0$ (which do not assume autocorrelation of ε_t).

The BNM_a and $BEPO_a$ statistics have substantially lower power than the BNM and BEPO statistics that use the other two information criteria. The power of BNM_h and $BEPO_h$ statistics is lower that the power of BNM_b and $BEPO_b$, but close to it. For sample sizes $N = 200, 400$ the choice of a specific information criterion does not make any substantial difference with respect to the level of power of the statistics.

Table 3.6 shows that MZ_a, MZ_t and MSB for $N = 50$ have substantially higher power than our statistics. For N higher that 100, our statistics appear to outperform the modified statistics derived by Ng and Perron (2001). The ADF statistic appears to have comparatively high power across all sample sizes considered. When an intercept and a trend are included in the model (Table 3.7), we observe that statistics MZ_a, MZ_t, MSB, P_T, MP_T, MZ_{aLS} and MZ_{a2} have extremely low power (smaller than the 5% size for alternatives close to H_0). The

ADF and ADF_{LS} statistics appear to have higher power compared to our statistics. For sample sizes higher than 100, BNM_b and $BEPO_b$ statistics have higher power than the ADF statistic.

Table 3.8 presents the results for size-adjusted power for the model including an intercept only, when there is negative autocorrelation ($\phi = -0.5$) in the error term. A first observation is that the power of our statistics is lower in comparison to the case of no autocorrelation (Table 3.6), especially for sample sizes 50 and 100. We also observe that the BNM_b, $BEPO_b$, BNM_h and $BEPO_h$ statistics appear to have higher power than BNM_0 and $BEPO_0$ for sample size $N = 50$. For this sample size the modified statistics perform better than our similar statistics. For sample size 100, we observe that BNM_b, $BEPO_b$, BNM_h and $BEPO_h$ have higher power than BNM_0 and $BEPO_0$ and the statistics proposed by Ng and Perron (2001) for alternatives far from the null $\rho = 1$. For alternatives $0.98 \leq \rho \leq 0.90$ we find that the ADF statistic has higher power that the rest of the statistics. For the same alternatives, BNM_0 and $BEPO_0$ have higher power. For higher sample sizes, our statistics have comparatively higher (in comparison to the Ng and Perron statistics) power for most alternatives.

Table 3.9 refers to the case of a model including an intercept and a trend in the presence of negative autocorrelation in the error term ($\phi = -0.5$). First of all, for sample size $N = 50$ we observe that all statistics suffer from the problem of very low power. We also observe that for most alternatives, BNM_0 and $BEPO_0$ exhibit higher power than our statistics for sample sizes $N = 50, 100$. Also the statistics derived by the procedure of Ng and Perron (2001) have substantially higher power than ours. For sample sizes higher than $N = 100$ our statistics appear to have higher power for most alternatives.

Tables 3.10 and 3.11 contain the finite sample power of the statistics when there is no autocorrelation of the error term, for alternatives farther than the ones investigated in Tables 3.6 and 3.7. The reason for this is to examine the possibility of power reversal. Table 3.10 corresponds to a model with an intercept only. We observe that the problem of power reversal is severe for statistics MZ_a, MZ_t, MSB, P_T, MP_T, MZ_{aLS} and MZ_{a2} for sample sizes $N = 100, 200$. For example for sample size $N = 100$ and alternative $\rho = 0.8$ the power of the MZ_a is 0.839 which is the highest among the values of power computed. Moving away from alternative $\rho = 0.8$ (to alternatives $\rho < 0.8$), power decreases gradually, reaching power 0.715 for alternative 0.1. ADF and ADF_{LS} statistics do not appear to have this problem.

Regarding the statistics derived in this Chapter, we can see that power reversal occurs for the BNM_a and $BEPO_a$ statistics. For the same case ($N = 100$) $BEPO_a$ statistic has power 0.958 for $\rho = 0.5$ and then gradually falls to 0.947 for $\rho = 0.1$. We consider the power reversal of $BEPO_a$ to be less serious than that occurring for MZ_a mainly because of the magnitude of the power reduction: 1.1% for $BEPO_a$ power reduction from alternative 0.5 to 0.1 is 1.1%, while the power reduction from alternative 0.8 to 0.1 for MZ_a is 14.8%. Additionally, we observe that our statistics have higher power in comparison to the other test statistics existing in the literature for alternatives far from the null ($\rho = 1$) and close to 0. This can be seen in Figure 3.4 which presents the power of the BNM_b, BNM_a, BNM_h, ADF, and MZ_a for the model including an intercept only. For sample size $N = 50$, the ADF statistic appears to have higher power that the BNM_b, BNM_a and BNM_h statistics. The MZ_a statistic has higher power for alternatives far from the null. For higher sample sizes the BNM_b, BNM_a and BNM_h perform better than ADF and MZ_a. In this figure one can see that the power function of MZ_a changes slope for sample sizes $N = 100, 200$.

The problem of power reversal becomes more apparent in the context of a model which includes an intercept and a trend. This case is presented for the same statistics in Figure 3.5. In this case, one can see that even for a sample size as high as $N = 400$, the MZ_a statistic has a decreasing power as the true value of ρ moves farther away from H_0. Table 3.11 presents the results for power in the absence of autocorrelation in the errors for all the statistics. For statistics MZ_a, MZ_t, MSB, P_T, MP_T, MZ_{aLS} and MZ_{a2} similar conclusions to those of Table 3.10 are drawn. Table 3.11 shows that the problem of power reversal occurs for statistic ADF as well, but not for ADF_{LS}. This problem appears for the BNM and BEPO statistics being less severe (much smaller power reduction as ρ moves farther away from the null).

The above discussion highlights the nature as well as the extent of the problem of power reversal for the statistics derived by Ng and Perron (2001). At this point, it is necessary to assess the performance of the test statistics which address this problem. These are the statistics derived by Perron and Qu (2007) and Seo (2006). We consider the ADF t-statistic and modified Phillips Perron statistic resulting from the Perron and Qu procedure (denoted as ADF_{PQ} and M_{PQ}) and the same statistics resulting from the Seo two-step procedure (denoted as ADF_S and M_S).

The purpose of our numerical study at this point, is to investigate two main questions:

- Do the modifications proposed by Perron and Qu (2007) and Seo (2006) solve the problem of power reversal?
- What is the effect of these modifications on size distortion as well as the level of power of the statistics, when compared with the statistics without the modification and the statistics derived in this Chapter?

We compare these procedures to the BNM and BEPO statistics, using asymptotic critical values. The BNM and BEPO statistics that use asymptotic critical values are denoted as $\overline{BNM}_i$ and $\overline{BEPO_i}$ with $i = a, b, h$.

Table 3.12 presents the empirical size of the statistics mentioned above in the context of a model including an intercept only. A first observation regarding the BNM and BEPO statistics is that when using asymptotic critical values (instead of exact), the size control in small samples is not always better when we use the AIC. It appears that the use of any of the information criteria results in the almost the same size distortion for values of ϕ close to -1. For values of ϕ closer to 0 the AIC appears to deliver higher size distortion than the BIC. For example, in Table

3.12 for sample size $N = 50$, $\phi = 0$, $\overline{BEPO_b}$ and $\overline{BEPO_a}$ deliver empirical size 0.058 and 0.104, respectively. The reason is the liberal nature of the AIC. In the absence of autocorrelation in the error term, AIC chooses higher MA order than the BIC. This difference is more evident in small samples. It is worth noting here that the $\overline{BNM_b}$ and $\overline{BEPO_b}$ statistics exhibit very close empirical size to BNM_b and $BEPO_b$ (see Tables 3.2 and 3.3). This means that BNM and BEPO statistics that use the BIC are robust to the use of exact or asymptotic critical values.

When comparing the Perron and Qu (2007) and Seo (2006) procedures we find that for a given statistic, the Perron and Qu procedure results in lower size distortion than the Seo procedure. Additionally, we see that the BNM and BEPO statistics generally perform better with respect to control over size than the ADF_S statistic. Comparing the statistics derived in this Chapter to the ADF_{PQ}, M_{PQ} and M_S statistics, we find that the latter three have lower size distortion. However, a problem regarding size mentioned above still remains for the modified Phillips Perron statistics (M_{PQ} and M_S): as sample size increases there is the possibility that empirical size deviates from that of the nominal. This can be seen by observing the empirical size of M_{PQ} for $\phi = -0.8$, which is 0.161, 0.042, 0.018, 0.017 for sample sizes $N = 50, 100, 200, 400$ respectively. However, the BNM and BEPO statistics using asymptotic critical values do not appear to suffer from this problem (their empirical size tends to the nominal one as sample size increases).

Table 3.13 presents the results regarding the empirical size of the statistics in the context of a model containing an intercept and a linear trend. A first observation is that size distortion is generally higher compared to the case of a model including an intercept only. We observe now that the BNM and BEPO statistics perform better than the ADF_S with respect to size distortion (as is the case in Table 3.12), but also BNM and BEPO perform better than M_S for sample size $N = 50$ and for values of ϕ close to $|1|$.

Figure 3.6 depicts the empirical size of the $\overline{BNM_b}$, $\overline{BEPO_b}$, ADF_{PQ}, M_{PQ}, ADF_S and M_S statistics in the context of a model with an intercept only. Figure 3.7 corresponds to a model including an intercept and a trend. These figures show that the ADF_{PQ}, and M_{PQ} statistics have low size distortion in comparison to the rest of the statistics. The $\overline{BNM_b}$, and $\overline{BEPO_b}$ statistics appear to have lower size distortion that the ADF_S and M_S statistics for values of ϕ close to -1.

Tables 3.14 and 3.15 present the performance of the statistics with respect to power for alternatives close to the null $(0.8, 0.82, ..., 0.98)$, for a model with an intercept only and a model with an intercept and a trend. We observe that the power of the statistics (for a given alternatives) is substantially lower in the case of a model including an intercept and a linear trend. In this case we can make some interesting observations, the first of which, is the detrimental effect of size being very low comparatively to its nominal level (rather than converging to it). Table 3.15 suggests that Phillips Perron statistic exhibit very low power in small samples such as $N = 50$. We see that the M_{PQ} statistic has a rejection (of the null hypothesis $H_0 : \rho = 1$) probability of 0.058 when the true value of the parameter $\rho = 0.80$, and when $\rho = 0.82$ its power (0.044) is lower than its nominal size (0.05).

The empirical size of the M_{PQ} statistic for the same case and sample size $N = 50$ is 0.009. The extremely low power of the statistic in this case, could be attributed to the fact that empirical size is far from its nominal value and close to zero.

Additionally, we observe that the ADF_S and M_S statistics appear to have higher power than ADF_{PQ} and M_{PQ} statistics. The BNM criterion appears to deliver statistics with higher power, for most alternatives and especially in small sample sizes, than the power of the statistics resulting from the BEPO criterion. Finally, we observe that the ADF_S appears to have the highest power compared to the rest of the statistics for alternatives close to the null (0.80 to 0.98− Tables 3.14 and 3.15).

Tables 3.16 and 3.17 present the results for power for alternatives farther from the null for the two models considered. The purpose of these tables is to examine the possibility of power reversal for the statistics. We decide to examine the case in which there is no autocorrelation in the error term ($\phi = 0$), as this is the case in which power reversal occurs. It is obvious that this problem is minimal (if at all existent) for the BNM and BEPO statistics, as their power increases as we move to alternatives farther from the null. However, this is not the case for ADF_{PQ}, ADF_S, M_{PQ}, and M_S. We observe that (for a given sample size) the power of the statistics does not increase monotonically as we move away from the null.

It is very interesting that although the Perron and Qu statistics are constructed in order to deal with the problem of power reversal, they reduce it but do not eliminate it. For example, in the context of a model including an intercept only and for sample size $N = 100$, the power of the M_{PQ} statistic is 0.739 for alternative $\rho = 0.6$, and 0.663 for $\rho = 0.1$. The problem is not solved as sample size increases. For sample size $N = 400$, we observe that the power reversal problem is obvious in both model specifications (presented in Tables 3.16 and 3.17) for statistics ADF_{PQ}, ADF_S, M_{PQ}, and M_S. Additionally, the power of statistics $\overline{BNM}_b$, $\overline{BEPO}_b$, ADF_{PQ}, ADF_S, M_{PQ}, and M_S is presented in Figures 3.8 and 3.9, for a model with an intercept only and a model with an intercept and a trend, respectively. It is obvious that the $\overline{BNM}_b$, and $\overline{BEPO}_b$ statistics have high power compared to the rest of the statistics (especially for sample size higher than 100) and that they do not suffer from the power reversal problem.

Observation of Tables 3.16 and 3.17 suggests another problem of the ADF_S, and M_S statistics. There are cases in which, for a given alternative, power decreases as sample size increases, i.e. more information (higher N) leads to worse inference (lower probability of rejecting the false null hypothesis). In Table 3.16, for example, we see that the power of M_S for alternative $\rho = 0.1$ is 0.9 for $N = 50$, 0.727 for $N = 100$, 0.635 for $N = 200$ and 0.623 for $N = 400$. The BNM and BEPO statistics do not present such behaviour: for given alternatives power increases as sample size increases.

Figures 3.8 and 3.9 present the power of the $\overline{BNM_b}$, $\overline{BEPO_b}$, ADF_{PQ}, M_{PQ}, ADF_S and M_S statistics for the two models considered in this Chapter. These show that for sample sizes higher than $N = 100$ the BNM and BEPO statistics perform comparatively very well. It is also obvious that the BNM and BEPO statistics do not suffer from the power reversal problem.

Tables 3.18 ($N = 50, 100$) and 3.19 ($N = 200, 400$) present the performance of the information criteria across different values of ϕ under H_0, for a model with an intercept only. This could help to explain the difference among our statistics with respect to control over size. As mentioned above, the BIC is the most conservative information criterion and AIC is the most liberal, while HQIC lies between the two other criteria. As a consequence, statistics that use the AIC have better control over size in the presence of negative MA parameters in comparison to statistics using the other criteria. For $N = 50$ and value $\phi = -0.8$ under H_0, the BIC chooses order 0 (no autocorrelation) 68.6% of the cases, AIC 39.4% and HQIC 52.7%. As sample size increases the performance of all information criteria is improved (they tend to choose the correct order) and for $N = 400$ and value $\phi = -0.8$, none of the information criteria chooses order 0 (i.e. all criteria suggest that there is autocorrelation in the error term). This is the reason why we do not observe substantial difference with respect to size distortion among our statistics for large samples. Information criteria behave similarly in the case of a model with an intercept and a trend.

Tables 3.20 and 3.21 present the performance of information criteria for models including an intercept only and an intercept and a trend respectively, across different values of alternatives in the case of no autocorrelation in the error term ($\phi = 0$). These tables explain the occurrence of the problem of power reversal for some of our statistics. We observe that the AIC performs worse with respect to identifying the right MA order as the true value of ρ moves farther away from the null. In Table 3.20, we see that for $N = 100$, the AIC chooses order 0 (the true order under the DGP) for the MA component 70.8% of the cases and for $\rho = 0.1$, 61.5%. For the same sample size, under the null, the BIC chooses order 0 for the MA component 95.6% of the cases and for $\rho = 0.1$, 93.9%. Table 3.21 shows that moving to a model with an intercept and a trend makes the problem of identifying the right order more serious for the AIC. For this model, and for sample size $N = 100$, the AIC chooses order 0 67.8% of the cases and for $\rho = 0.1$, choice of zero order falls to 55.5%. Under H_0, the BIC chooses order 0 94.8% of the cases and for $\rho = 0.1$, the relative frequency is 92%.

3.7. Conclusion

In this Chapter we derive asymptotically similar statistics for testing the unit root hypothesis in the presence of autocorrelated errors. Based on the BNM and BEPO optimality criteria proposed by Forchini and Marsh (2000), we derive test statistics that take into consideration possible autocorrelation in the error term. We consider our testing procedure to be feasible with respect to two aspects. The first involves the use of information criteria (BIC, AIC and HQIC) for the choice of the order of autocorrelation. The second includes the estimation of the parameters of the chosen model. Limiting distributions for the test statistics are provided which enable us to use asymptotic critical values for high sample sizes (over $N =$

100). In order to assess the finite sample performance of our statistics under different specifications, we perform an extensive simulation study.

We believe that we successfully generalize the statistics of Forchini and Marsh as we substantially improve the size control of the statistics in the presence of autocorrelation, without any significant power loss even in the case of no autocorrelation in the error term.

Additionally, we compare our statistics with a variety of other statistics existing in the literature (mainly those in Ng and Perron, 2001). We find that for a small sample size (such as $N = 50$) the other statistics could possibly have so high level of size distortion, that would make inference drawn by them highly unreliable. Our test statistics perform much better with respect to control over size. For higher sample sizes, our statistics perform comparatively worse to the Ng and Perron statistics, but size distortion appears to fall substantially as sample size increases. With respect to finite sample power, our statistics achieve higher power for most alternatives apart from those close to the null hypothesis. Finally, the asymptotically similar BNM and BEPO statistics (using exact or asymptotic critical values) do not seem to suffer seriously from the problem of power reversal.

Moreover, we compare our statistics to the statistics resulting from the Perron and Qu (2007) and Seo (2006) procedures. A striking observation of the numerical study on these procedure is the severity of the power reversal problem. In light of the results provided by Seo (2006), Perron and Qu (2007) modify the Ng and Perron (2001) statistics in order to solve the power reversal problem. However, even if there is improvement, the problem is still evident. Additionally, the statistics proposed by Seo (2006) suffer from the same problem.

The above observation leads to a methodological aspect highlighted in this Chapter regarding the examination of power properties. More specifically, the power of the statistics should be checked for alternatives (comparatively) far from the null. Investigation of such alternatives could reveal possible power reversal problems.

Another methodological issue discussed in this Chapter is the caution that should be exercised for cases in which empirical size tends to zero rather than to its nominal value. *Prima facie* this could appear as a good property, as size is the probability of an error (rejecting the null hypothesis when it is true) which we want to keep lower than a certain level (nominal size). However, having size tending to zero comes at the expense of low power. Presenting size-adjusted power could hide this problem. That is the reason that we believe the size of a *well-behaved* statistic should tend to its nominal value as sample size increases, otherwise we face the possibility of adverse effects on the ability of the statistic to reject the null hypothesis when this hypothesis is not true.

We examine the finite sample properties of the statistics derived in this Chapter for exact and asymptotic critical values. When exact critical values are used, we observe that the optimality criteria used (BNM and BEPO) deliver statistics that have very similar empirical size and power in finite samples. However, what differentiates the finite sample properties of our statistics, is the use of the information criterion for the determination of the order of the MA component. The use of AIC delivers the best results with respect to size control, but also has the lowest power and for some sample sizes the problem of power reversal occurs. The BIC provides the best results with respect to power, but the worst for controlling size in small samples. The HQIC appears to lie in between the other criteria mentioned, delivering test statistics with power close to BNM_b and $BEPO_b$, and size distortion not much higher than the one of BNM_a and $BEPO_a$. We suggest the use of the HQIC, because the BNM_h and $BEPO_h$ statistics appear to have, in comparison to the other asymptotically similar statistics, low size distortion, high power and not significant (if any) power reduction for alternatives far from the null.

When asymptotic critical values are used, the main findings of the numerical analysis do not change. One difference though is that the empirical size of the statistics using the BIC tends faster to its nominal value as sample size increases for $\phi = 0$. For values of ϕ close to -1, AIC gives better results with respect to control over size.

In concluding the discussion about the comparison of the statistics derived in this Chapter to the other unit root test statistics in the literature, we believe that the main strength of the statistics is their robustness. More specifically they appear to have comparatively good control over size across different values of ϕ and high power across a wide range of alternatives. Additionally, they improve with respect to size control and power as sample size increases and do not suffer from the power reversal problem.

3.8. Technical Appendix and Proofs

Proposition A1. *The lag matrix* $L^{(i)}$ *commutes with any other lag matrix of different or same order* $L^{(j)}$ *and*

$$\begin{aligned}
K_\phi T_\rho &= T_\rho K_\phi, \\
T_1^{-1} K_\phi^{-1} &= K_\phi^{-1} T_1^{-1}, \\
T_1' K_\phi' &= K_\phi' T_1', \\
\left(T_1^{-1}\right)' \left(K_\phi^{-1}\right)' &= \left(K_\phi^{-1}\right)' \left(T_1^{-1}\right)', \\
K_\phi^{-1} T_\rho &= T_\rho K_\phi^{-1}, \\
\left(K_\phi^{-1}\right)' T_\rho' &= T_\rho' \left(K_\phi^{-1}\right)',
\end{aligned}$$

given that K_ϕ *and* T_ρ *are invertible.*

Proof. Lag matrix $L^{(i)}$ commutes with any other lag matrix of the same or different order $L^{(j)}$ and:

$$L^{(i)} L^{(j)} = L^{(j)} L^{(i)} = \begin{cases} L^{(i+j)}, & \text{for } i + j \leq N - 1 \\ 0, & \text{for } i + j > N - 1. \end{cases} \tag{3.24}$$

Noting the definitions in (3.5) and (3.6) and the commutative property of lag matrix $L^{(i)}$ (3.24) we have:

$$\begin{aligned}
K_\phi T_\rho &= \left(I_N + \sum_{i=1}^{q} \phi_i L^{(i)}\right)\left(I_N - \rho L^{(1)}\right) \\
&= I_N - \rho L^{(1)} + \sum_{i=1}^{q} \phi_i L^{(i)} - \left(\sum_{i=1}^{q} \phi_i L^{(i)}\right) \rho L^{(1)} \\
&= I_N - \rho L^{(1)} + \sum_{i=1}^{q} \phi_i L^{(i)} - \rho \sum_{i=1}^{q} \phi_i L^{(i)} L^{(1)} \\
&= I_N - \rho L^{(1)} + \sum_{i=1}^{q} \phi_i L^{(i)} - \rho L^{(1)} \sum_{i=1}^{q} \phi_i L^{(i)} \\
&= I_N - \rho L^{(1)} + \left(I_N - \rho L^{(1)}\right) \sum_{i=1}^{q} \phi_i L^{(i)} \\
&= \left(I_N - \rho L^{(1)}\right)\left(I_N + \sum_{i=1}^{q} \phi_i L^{(i)}\right) = T_\rho K_\phi. \qquad (3.25)
\end{aligned}$$

Equation (3.25) means that K_ϕ commutes with T_ρ (and with T_1 which is a special case of T_ρ). Given that K_ϕ and T_ρ are nonsingular matrices, we can easily show that their respective inverse and transpose matrices commute with each other as well:

$$K_\phi T_\rho = T_\rho K_\phi \Leftrightarrow (K_\phi T_\rho)^{-1} = (T_\rho K_\phi)^{-1} \Leftrightarrow T_\rho^{-1} K_\phi^{-1} = K_\phi^{-1} T_\rho^{-1}, \qquad (3.26)$$

$$K_\phi T_\rho = T_\rho K_\phi \Leftrightarrow (K_\phi T_\rho)' = (T_\rho K_\phi)' \Leftrightarrow T_\rho' K_\phi' = K_\phi' T_\rho', \qquad (3.27)$$

and combining (3.26) and (3.27) we get

$$\left(T_1^{-1}\right)'\left(K_\phi^{-1}\right)' = \left(K_\phi^{-1}\right)'\left(T_1^{-1}\right)'. \tag{3.28}$$

Finally, using (3.25) we show that T_ρ commutes with K_ϕ^{-1}

$$K_\phi T_\rho = T_\rho K_\phi \Rightarrow T_\rho = K_\phi^{-1} T_\rho K_\phi \Rightarrow T_\rho K_\phi^{-1} = K_\phi^{-1} T_\rho, \tag{3.29}$$

and transposing both sides of (3.25) we can show that $\left(K_\phi^{-1}\right)' T_\rho' = T_\rho' \left(K_\phi^{-1}\right)'$.

Proposition A2. *Let* $S = T_1^{-1}\varepsilon$ *and* $\sigma^2 = E\left(\varepsilon_1^2\right)$. *Under the assumptions of Theorem 3.3 with* X satisfying (3.22), *the following limit theory applies under the null hypothesis* $H_0 : \rho = 1$ *as* $N \to \infty$*:*

(i) $N^{-1}S'\varepsilon \Rightarrow \frac{1}{2}\sigma^2\left[W^2(1) + 1\right]$

(ii) $N^{-1}S'P_Z\varepsilon \Rightarrow \sigma^2 W(1)\int_0^1 W(r)dr$

(iii) $N^{-2}S'T_1^{-1}P_Z\varepsilon \Rightarrow \sigma^2 W(1)\int_0^1 rW(r)dr$

(iv) $N^{-2}\left(T_1^{-1}P_Z\varepsilon\right)' T_1^{-1}P_Z\varepsilon \Rightarrow \frac{1}{3}\sigma^2 W^2(1)$

(v) $N^{-1}\left(T_1^{-1}P_Z\varepsilon\right)' P_Z\varepsilon \Rightarrow \frac{1}{2}\sigma^2 W^2(1)$

(vi) $N^{-1}\left(T_1^{-1}P_Z\varepsilon\right)'\varepsilon \Rightarrow \sigma^2 W(1)\left(W(1) - \int_0^1 W(r)dr\right)$

(vii) $N^{-1}\nu'\nu \to_p \sigma^2$

For X *satisfying (3.23) parts (i) and (vii) continue to apply and:*

(viii) $N^{-1}S'P_Z\varepsilon$, $N^{-1}\left(T_1^{-1}P_Z\varepsilon\right)' P_Z\varepsilon$ and $N^{-2}\left(T_1^{-1}P_Z\varepsilon\right)' T_1^{-1}P_Z\varepsilon$ have order $O_p(N^{-1})$ and $N^{-2}S'T_1^{-1}P_Z\varepsilon$, $N^{-1}\left(T_1^{-1}P_Z\varepsilon\right)'\varepsilon$ have order $O_p(N^{-1/2})$ as $N \to \infty$.

where $W(\cdot)$ denotes standard Brownian motion on $D[0,1]$.

Proof. By definition of the matrix T_1^{-1}, S_t is a unit root process with i.i.d. innovations ε_t. Also, using the particular form of the matrix X of deterministics, it is easy to obtain the following identities:

$$P_Z\varepsilon = \frac{1}{N-1}\left[(N-1)\varepsilon_1, S_N-\varepsilon_1, ..., S_N-\varepsilon_1\right]'$$

and

$$T_1^{-1}P_Z\varepsilon = \frac{1}{N-1}\left[(N-1)\varepsilon_1, S_{N-1}+(N-1)\varepsilon_1, ..., (N-1)S_{N-1}+(N-1)\varepsilon_1\right]'.$$

In what follows, we make use of standard unit root asymptotics, see e.g. Phillips (1987b) and Phillips and Perron (1988).

For part (i), we have

$$\begin{aligned} N^{-1}S'\varepsilon &= N^{-1}\sum_{i=1}^{N} S_i\varepsilon_i = N^{-1}\left(\sum_{i=1}^{N} S_{i-1}\varepsilon_i + \sum_{i=1}^{N}\varepsilon_i^2\right) \\ &= N^{-1}\sum_{i=1}^{N} S_{i-1}\varepsilon_i + N^{-1}\sum_{i=1}^{N}\varepsilon_i^2 \\ &\Rightarrow \frac{1}{2}\sigma^2\left\{[W(1)]^2-1\right\}+\sigma^2 \\ &= \frac{1}{2}\sigma^2\left[W^2(1)+1\right]. \end{aligned}$$

For part (ii),

$$\begin{aligned}
\frac{1}{N}S'P_Z\varepsilon &= \frac{1}{N-1}\left[S_1\left(N-1\right)\varepsilon_1+\sum_{i=2}^{N}S_i\left(S_N-\varepsilon_1\right)\right] \\
&= \frac{1}{N\left(N-1\right)}S_N\sum_{i=2}^{N}S_i+O_p\left(\frac{1}{N^2}\sum_{i=2}^{N}S_i\right) \\
&= \frac{1}{N^{1/2}}S_N\frac{1}{N^{3/2}}\sum_{i=2}^{N}S_i+O_p(N^{-1/2}) \\
&\Rightarrow \sigma^2W(1)\int_0^1 W(r)dr.
\end{aligned}$$

For part (iii),

$$\begin{aligned}
\frac{1}{N^2}S'T_1^{-1}P_Z\varepsilon &= \frac{1}{N^2}\frac{1}{N-1}\sum_{i=1}^{N}\left\{S_i\left[\left(i-1\right)S_{N-1}+\left(N-1\right)\varepsilon_1\right]\right\} \\
&= \frac{1}{N^2}\frac{1}{N-1}S_{N-1}\sum_{i=1}^{N}S_i i+O_p\left(\frac{1}{N^2}\sum_{i=1}^{N}S_i\right) \\
&= \frac{1}{N^{1/2}}S_{N-1}\frac{1}{N^{5/2}}\sum_{i=1}^{N}S_i i+O_p\left(N^{-1/2}\right) \\
&\Rightarrow \sigma^2W(1)\int_0^1 rW(r)dr.
\end{aligned}$$

For part (iv),

$$\begin{aligned}
\frac{1}{N^2}\left(T_1^{-1}P_Z\varepsilon\right)' T_1^{-1}P_Z\varepsilon &= \frac{1}{N^2}\left(\frac{1}{N-1}\right)^2 \sum_{i=1}^{N}\left[(i-1)\,S_{N-1}+(N-1)\,\varepsilon_1\right]^2 \\
&= \frac{S_{N-1}^2}{(N-1)^2}\frac{1}{N^2}\sum_{i=1}^{N}(i-1)^2 + O_p\left(\frac{1}{N}S_{N-1}\right) \\
&= [1+o\,(1)]\,\frac{S_{N-1}^2}{3N} + O_p\left(N^{-1/2}\right) \\
&\Rightarrow \frac{1}{3}\sigma^2 W^2(1).
\end{aligned}$$

For part (v),

$$\begin{aligned}
\frac{1}{N}\left(T_1^{-1}P_Z\varepsilon\right)' P_Z\varepsilon &= \frac{1}{N}\left(\frac{1}{N-1}\right)^2 \left\{(N-1)^2\,\varepsilon_1^2 + \sum_{i=1}^{N-1}\left[\left(iS_{N-1}+(N-1)\,\varepsilon_1\right)\left(S_N-\varepsilon_1\right)\right]\right\} \\
&= \frac{1}{N\,(N-1)^2}S_{N-1}S_N\sum_{i=1}^{N-1} i + O_p\left(\frac{1}{N^3}S_{N-1}\sum_{i=1}^{N-1} i\right) \\
&= [1+o\,(1)]\,\frac{1}{2}\frac{S_{N-1}}{N^{1/2}}\frac{S_N}{N^{1/2}} + O_p\left(N^{-1/2}\right) \\
&\Rightarrow \frac{1}{2}\sigma^2 W^2(1)
\end{aligned}$$

For part (vi),

$$\begin{aligned}
\frac{1}{N}\left(T_1^{-1}P_Z\varepsilon\right)'\varepsilon &= \frac{1}{N}\left\{\frac{1}{N-1}S_{N-1}\sum_{i=1}^{N} i\varepsilon_i - \frac{1}{N-1}S_{N-1}S_N + \varepsilon_1 S_N\right\} \\
&= [1+o\,(1)]\,\frac{S_{N-1}}{N^{1/2}}\frac{1}{N^{3/2}}\sum_{i=1}^{N} i\varepsilon_i + O_p\left(N^{-1/2}\right) \\
&\Rightarrow \sigma W(1)\left(\sigma W(1) - \sigma\int_0^1 W(r)dr\right).
\end{aligned}$$

For part (vii), recall that, under H_0, $Z = K_{\hat{\phi}}^{-1} T_1 X$ and $T_1 u = K_\phi \varepsilon$ which gives

$$\begin{aligned} \nu &= M_Z K_{\hat{\phi}}^{-1} T_1 \left(X\beta + u \right) = M_Z K_{\hat{\phi}}^{-1} T_1 u \\ &= M_Z K_{\hat{\phi}}^{-1} K_\phi \varepsilon = \left[I + o_p\left(1\right) \right] M_Z \varepsilon \end{aligned}$$

using the fact that $\hat{\phi} - \phi = o_p\left(1\right)$. Therefore, since

$$\varepsilon' P_Z \varepsilon = \varepsilon_1^2 + \frac{1}{N-1} \left(S_N - \varepsilon_1 \right)^2 = O_p\left(1\right),$$

the weak law of large numbers yields

$$\begin{aligned} \frac{1}{N} \nu' \nu &= \left[I + o_p\left(1\right) \right] \frac{1}{N} \varepsilon' M_Z \varepsilon \\ &= \left[I + o_p\left(1\right) \right] \left\{ \frac{1}{N} \varepsilon' \varepsilon + O_p\left(N^{-1}\right) \right\} \to_p \sigma^2. \end{aligned}$$

For part (viii) P_Z corresponds to X including a constant term only which gives the following results,

$$\begin{aligned} S' P_Z \varepsilon &= \left(T_1^{-1} P_Z \varepsilon \right)' P_Z \varepsilon = \varepsilon_1^2 = O_p(1), \\ \left(T_1^{-1} P_Z \varepsilon \right)' T_1^{-1} P_Z \varepsilon &= N \varepsilon_1^2 = O_p(N), \\ S' T_1^{-1} P_Z \varepsilon &= \varepsilon_1 \sum_{i=1}^{N} S_i = O_p(N^{3/2}), \\ \left(T_1^{-1} P_Z \varepsilon \right)' \varepsilon &= \varepsilon_1 \sum_{i=1}^{N} \varepsilon_i = \varepsilon_1 S_N = O_p(N^{1/2}), \end{aligned}$$

A direct result from the above is that $N^{-1}S'P_Z\varepsilon$, $N^{-1}\left(T_1^{-1}P_Z\varepsilon\right)'P_Z\varepsilon$ and $N^{-2}\left(T_1^{-1}P_Z\varepsilon\right)'T_1^{-1}P_Z\varepsilon$ have order $O_p(N^{-1})$ and $N^{-2}S'T_1^{-1}P_Z\varepsilon$, $N^{-1}\left(T_1^{-1}P_Z\varepsilon\right)'\varepsilon$ have order $O_p(N^{-1/2})$ as $N \to \infty$.

Proof of Lemma 3.1. Using the commutation results given in Proposition A1 we get

$$\begin{aligned}
\Sigma_{\rho,\phi} &= K_\phi^{-1}T_1T_\rho^{-1}K_\phi K_\phi'\left(T_\rho^{-1}\right)'T_1'\left(K_\phi^{-1}\right)' = T_1K_\phi^{-1}K_\phi T_\rho^{-1}\left(T_\rho^{-1}\right)'K_\phi'\left(K_\phi^{-1}\right)'T_1' \\
&= T_1T_\rho^{-1}\left(T_\rho^{-1}\right)'T_1'.
\end{aligned}$$

Proof of Theorem 3.1. The most powerful similar test of size α is given by (3.13) which can be rewritten as:

$$\frac{y'T_1'\left(K_\phi^{-1}\right)'C\left(C'\Sigma_\rho C\right)^{-1}C'K_\phi^{-1}T_1y}{y'T_1'\left(K_\phi^{-1}\right)'M_ZK_\phi^{-1}T_1y} < k_\alpha$$

Lemma 3 of Forchini and Marsh (2000) shows that the matrix

$$Q = C'B^{-1}C - \left(C'BC\right)^{-1}$$

is positive semi-definite. Applying this in our case gives the following result:

$$\begin{aligned}
\frac{y'T_1'\left(K_\phi^{-1}\right)'C\left(C'\Sigma_\rho C\right)^{-1}C'K_\phi^{-1}T_1y}{y'T_1'\left(K_\phi^{-1}\right)'M_ZK_\phi^{-1}T_1y} &\leq \frac{y'T_1'\left(K_\phi^{-1}\right)'CC'\Sigma_\rho^{-1}CC'K_\phi^{-1}T_1y}{y'T_1'\left(K_\phi^{-1}\right)'M_ZK_\phi^{-1}T_1y} \\
\frac{y'T_1'\left(K_\phi^{-1}\right)'C\left(C'\Sigma_\rho C\right)^{-1}C'K_\phi^{-1}T_1y}{y'T_1'\left(K_\phi^{-1}\right)'M_ZK_\phi^{-1}T_1y} &\leq \frac{\nu'\Sigma_\rho^{-1}\nu}{\nu'\nu},
\end{aligned}$$

where ν is defined above. So (3.13) is bounded above by the ratio of quadratic forms in ν. Inverting Σ_ρ and expressing T_ρ as $T_\rho = I_N - \rho L^{(1)}$:

$$\Sigma_\rho^{-1} = \left[T_1 T_\rho^{-1} \left(T_\rho^{-1}\right)' T_1'\right]^{-1} = \left(T_1^{-1}\right)' T_\rho' T_\rho T_1^{-1} =$$
$$\left(T_1^{-1}\right)' \left(I_N - \rho L^{(1)}\right)' \left(I_N - \rho L^{(1)}\right) T_1^{-1} =$$
$$\left(T_1^{-1}\right)' \left(I_N - \rho L^{(1)}\right) T_1^{-1} - \rho \left(T_1^{-1}\right)' L^{(1)\prime} \left(I_N - \rho L^{(1)}\right) T_1^{-1} =$$
$$\left(T_1^{-1}\right)' T_1^{-1} - \rho \left(T_1^{-1}\right)' L^{(1)} T_1^{-1} - \rho \left(T_1^{-1}\right)' L^{(1)\prime} T_1^{-1} + \rho^2 \left(T_1^{-1}\right)' L^{(1)\prime} L^{(1)} T_1^{-1}. \tag{3.30}$$

From equation (3.30) and the definition of the matrix $\Psi(\nu)$ we obtain:

$$\frac{\nu' \Sigma_\rho^{-1} \nu}{\nu' \nu} = \begin{pmatrix} 1 & -\rho \end{pmatrix} \Psi(\nu) \begin{pmatrix} 1 \\ -\rho \end{pmatrix} \tag{3.31}$$

So a sufficient condition for (3.13) to hold is that the positive definite matrix $\Psi(\nu)$ is small with respect to some norm. We can find statistics such that $\Pr\{\|\Psi(\nu)\| < k_\alpha | H_0\} = a$.

Proof of Theorem 3.2. The first BEPO criterion is:

$$\frac{l(\rho^*)' \Psi(y) l(\rho^*)}{y' \Omega^{-1}(\rho_0) y} < k_a \tag{3.32}$$

where k_α is such that the size of the test is α and ρ^* is the value of ρ which minimizes (3.31). We differentiate (3.31) with respect to parameter ρ and set it equal to zero. From equations (3.18) and (3.31) we get:

$$\begin{aligned}
\begin{pmatrix} 1 & -\rho \end{pmatrix} \Psi(\nu) \begin{pmatrix} 1 \\ -\rho \end{pmatrix} &= \left(\rho^2 \nu' \Psi_{22} \nu - 2\rho \nu' \Psi_{12} \nu + \nu' \Psi_{11} \nu\right). \\
\frac{1}{\nu'\nu} \frac{\partial \left(\rho^{*2} \nu' \Psi_{22} \nu - 2\rho^* \nu' \Psi_{12} \nu + \nu' \Psi_{11} \nu\right)}{\partial \rho^*} &= 0 \Rightarrow \\
\frac{2}{\nu'\nu} \left(\rho^* \psi_{22} - \psi_{12}\right) &= 0 \Rightarrow \rho^* = \frac{\psi_{12}}{\psi_{22}}. \qquad (3.33)
\end{aligned}$$

Combining condition (3.32) with (3.33) and values given by (3.16) and (3.17) we get the BEPO statistic. Also we need to note that $\psi_{22} \geq 0$ since Ψ_{22} is a positive semi-definite matrix, so $\frac{\partial^2 \left(\rho^2 \psi_{22} - 2\rho \psi_{12} + \psi_{11}\right)}{\partial \rho^2} \geq 0$.

The theorem is proved by substituting (3.33) and in (3.3).

Proof of Theorem 3.3. We make repeated use of the limit theory established in Proposition A2. For notational simplicity, define

$$\psi_{11} = \nu' \Psi_{11} \nu, \; \psi_{22} = \nu' \Psi_{22} \nu \text{ and } \psi_{12} = \nu' \Psi_{12} \nu$$

and note that

$$\begin{aligned}
\psi_{22} &= \psi_{11} - 2\left(T_1^{-1}\nu\right)' \nu + \nu'\nu \\
&= \psi_{11} - 2\left[S'\varepsilon - S' P_Z \varepsilon + \left(T_1^{-1} P_Z \varepsilon\right)' P_Z \varepsilon\right] \qquad (3.34) \\
&\quad -2\left(T_1^{-1} P_Z \varepsilon\right)' \varepsilon + \nu'\nu
\end{aligned}$$

and

$$\psi_{12} = \psi_{11} - S'\varepsilon + S'P_Z\varepsilon - \left(T_1^{-1}P_Z\varepsilon\right)' P_Z\varepsilon. \tag{3.35}$$

For part (i), it is clear that from Proposition A2 and (3.34) and (3.35) we obtain that $\psi_{22} = \psi_{11} + O_p(N)$ and $\psi_{12} = \psi_{11} + O_p(N)$. Now by Proposition A2,

$$\begin{aligned} \frac{1}{N^2}\psi_{11} &= \frac{1}{N^2}S'S - \frac{2}{N^2}S'T_1^{-1}P_Z\varepsilon + \frac{1}{N^2}\left(T_1^{-1}P_Z\varepsilon\right)' T_1^{-1}P_Z\varepsilon \\ &\Rightarrow \sigma^2\left\{\int_0^1 W^2(r)dr - 2W(1)\int_0^1 rW(r)dr + \frac{1}{3}W^2(1)\right\}. \end{aligned} \tag{3.36}$$

The BNM test statistic is given by

$$\begin{aligned} \frac{1}{N}\left\|\Psi(\nu)\right\| &= \frac{1}{N^{-1}\nu'\nu}\left\|\frac{1}{N^2}\begin{bmatrix}\psi_{11} & \psi_{12}\\ \psi_{12} & \psi_{22}\end{bmatrix}\right\| \\ &= \frac{1}{N^{-1}\nu'\nu}\left\|\frac{1}{N^2}\begin{bmatrix}\psi_{11} & \psi_{12}\\ \psi_{12} & \psi_{22}\end{bmatrix}\right\| \\ &= \frac{1}{N^{-1}\nu'\nu}\left\|\frac{\psi_{11}}{N^2}\begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix}\right\| + O_p\left(N^{-1}\right) \\ &\Rightarrow \left\|\begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix}\right\|\left\{\int_0^1 W^2(r)dr - 2W(1)\int_0^1 rW(r)dr + \frac{1}{3}W^2(1)\right\} \end{aligned}$$

and the result follows from Proposition A2(vi) and (3.36).

Part (iii) corresponds to the case of a constant only included in the model. Proposition A2(viii) applies here and we get

$$\frac{1}{N^2}\psi_{11} = \frac{1}{N^2}S'S + O_p(N^{-1/2}) \Rightarrow \sigma^2 \int_0^1 W^2(r)dr. \tag{3.37}$$

The above result in conjunction with Proposition A2(vii) gives us

$$\frac{1}{N}\left\|\Psi(\nu)\right\| = \frac{1}{N^{-1}\nu'\nu}\left\|\frac{1}{N^2}\begin{bmatrix} \psi_{11} & \psi_{12} \\ \psi_{12} & \psi_{22} \end{bmatrix}\right\| \Rightarrow \left\|\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\right\| \int_0^1 W^2(r)dr.$$

For part (ii):

$$\begin{aligned}
\frac{1}{N}\left(\psi_{12}-\psi_{22}\right) &= \frac{1}{N}\left(S'\varepsilon - S'P_Z\varepsilon - \left(T_1^{-1}P_Z\varepsilon\right)'\varepsilon + \left(T_1^{-1}P_Z\varepsilon\right)' P_Z\varepsilon - \nu'\nu\right) \\
& \xrightarrow{L} \frac{1}{2}\sigma^2\left[W^2(1)+1\right] - \sigma^2 W(1)\int_0^1 W(r)dr \\
& -\sigma^2 W(1)\left(W(1) - \int_0^1 W(r)dr\right) + \frac{1}{2}\sigma^2 W^2(1) - \sigma^2 \\
&= \sigma^2\left\{\begin{array}{c} \frac{1}{2}\left[W^2(1)+1\right] - W(1)\int_0^1 W(r)dr \\ -W(1)\left(W(1)-\int_0^1 W(r)dr\right) + \frac{1}{2}W^2(1) - 1 \end{array}\right\} \\
&= \sigma^2\left\{\begin{array}{c} \frac{1}{2}W^2(1) + \frac{1}{2} - W(1)\int_0^1 W(r)dr \\ -W^2(1) + W(1)\int_0^1 W(r)dr + \frac{1}{2}W^2(1) - 1 \end{array}\right\} = -\frac{1}{2}\sigma^2.
\end{aligned}$$

As before, when a constant and trend are included in the model $N^{-2}\psi_{22} = N^{-2}\psi_{11} + O_p(N^{-1})$. Combining the above results and the one in (3.36), we get the asymptotic distribution of BEPO statistic which is given by

$$\begin{aligned} BEPO &= N\left|\frac{\psi_{12}-\psi_{22}}{\psi_{22}}\right| = \left|\frac{N^{-1}(\psi_{12}-\psi_{22})}{N^{-2}\psi_{22}}\right| \\ &\xrightarrow{L} \left|\frac{-\frac{1}{2}\sigma^2}{\sigma^2\left[\int_0^1 W^2(r)dr - 2W(1)\int_0^1 rW(r)dr + \frac{1}{3}W^2(1)\right]}\right| \\ &= \left|\frac{-\frac{1}{2}}{\int_0^1 W^2(r)dr - 2W(1)\int_0^1 rW(r)dr + \frac{1}{3}W^2(1)}\right| \\ &= \frac{1}{2}\frac{1}{\left|\int_0^1 W^2(r)dr - 2W(1)\int_0^1 rW(r)dr + \frac{1}{3}W^2(1)\right|}. \end{aligned}$$

For part (iv) of the theorem X satisfies (3.23). We use results from Proposition A2(viii) and we get

$$\begin{aligned} \frac{1}{N}(\psi_{12}-\psi_{22}) &= \frac{1}{N}\left(S'\varepsilon - S'P_Z\varepsilon - \left(T_1^{-1}P_Z\varepsilon\right)'\varepsilon + \left(T_1^{-1}P_Z\varepsilon\right)'P_Z\varepsilon - \nu'\nu\right) \\ &= \frac{1}{N}(S'\varepsilon - \nu'\nu) + O_p\left(N^{-1/2}\right) \\ \xrightarrow{L} \frac{1}{2}\sigma^2\left[W^2(1)+1\right] - \sigma^2 &= \frac{1}{2}\sigma^2\left[W^2(1)-1\right]. \end{aligned}$$

Using the above result and (3.37) we get

$$\begin{aligned} BEPO &= N\left|\frac{\psi_{12}-\psi_{22}}{\psi_{22}}\right| = \left|\frac{N^{-1}(\psi_{12}-\psi_{22})}{N^{-2}\psi_{22}}\right| \\ \xrightarrow{L} \left|\frac{\frac{1}{2}\sigma^2\left[W^2(1)-1\right]}{\sigma^2\int_0^1 W^2(r)dr}\right| &= \frac{1}{2}\left|\frac{W^2(1)-1}{\int_0^1 W^2(r)dr}\right|. \end{aligned}$$

3.9. Tables and Figures

Table 3.1. Asymptotic critical values.

	BNM			BEPO		
Percentile	1%	5%	10%	90%	95%	99%
	Case: Intercept only					
	0.069173	0.113954	0.15428	5.665921	7.98514	13.69566
	Case: Intercept and Trend					
	0.049464	0.07287	0.091695	10.90568	13.72295	20.21404

Table 3.2. Empirical size of the tests for model with an intercept only.

N	ϕ	BNM_0	$BEPO_0$	BNM_b	$BEPO_b$	BNM_a	$BEPO_a$	BNM_h	$BEPO_h$	MZ_a	MZ_t	MSB	P_t	MP_t	ADF	MZ_{aLS}	ADF_{LS}	MZ_{a2}
50	−0.8	0.669	0.673	0.496	0.501	0.367	0.367	0.433	0.437	0.992	0.988	0.992	0.931	0.986	0.997	0.975	0.992	0.976
	−0.7	0.559	0.563	0.370	0.367	0.253	0.251	0.298	0.305	0.932	0.916	0.930	0.807	0.909	0.957	0.831	0.918	0.903
	−0.6	0.464	0.466	0.261	0.260	0.179	0.176	0.196	0.196	0.791	0.762	0.783	0.637	0.751	0.842	0.599	0.759	0.751
	−0.5	0.364	0.366	0.215	0.214	0.149	0.148	0.163	0.164	0.605	0.573	0.593	0.454	0.564	0.680	0.389	0.568	0.569
	−0.4	0.255	0.259	0.175	0.170	0.112	0.115	0.127	0.130	0.435	0.410	0.425	0.319	0.399	0.520	0.220	0.389	0.407
	−0.3	0.155	0.162	0.152	0.154	0.081	0.084	0.096	0.102	0.292	0.271	0.281	0.213	0.267	0.364	0.111	0.247	0.271
	−0.2	0.135	0.135	0.104	0.101	0.077	0.076	0.077	0.076	0.181	0.164	0.172	0.128	0.162	0.237	0.052	0.157	0.167
	−0.1	0.068	0.070	0.076	0.073	0.064	0.064	0.060	0.063	0.104	0.095	0.097	0.075	0.095	0.149	0.022	0.098	0.097
	0	0.050	0.054	0.049	0.045	0.047	0.046	0.049	0.050	0.062	0.056	0.058	0.044	0.056	0.096	0.008	0.057	0.057
	0.1	0.023	0.024	0.049	0.045	0.047	0.045	0.046	0.047	0.041	0.038	0.038	0.031	0.037	0.068	0.003	0.046	0.038
	0.2	0.025	0.027	0.052	0.048	0.048	0.047	0.047	0.048	0.025	0.023	0.025	0.019	0.023	0.045	0.001	0.037	0.023
	0.3	0.018	0.019	0.047	0.047	0.056	0.056	0.045	0.046	0.016	0.014	0.015	0.011	0.013	0.028	0.001	0.031	0.015
	0.4	0.010	0.010	0.057	0.057	0.050	0.050	0.050	0.048	0.011	0.009	0.010	0.007	0.009	0.021	0.000	0.029	0.010
	0.5	0.011	0.010	0.064	0.061	0.052	0.052	0.044	0.045	0.009	0.009	0.007	0.007	0.008	0.018	0.000	0.030	0.008
	0.6	0.007	0.007	0.062	0.061	0.039	0.038	0.039	0.041	0.008	0.007	0.006	0.006	0.007	0.015	0.000	0.029	0.007
	0.7	0.008	0.009	0.053	0.053	0.039	0.039	0.035	0.039	0.005	0.005	0.005	0.005	0.005	0.014	0.000	0.027	0.005
	0.8	0.009	0.008	0.052	0.050	0.039	0.040	0.048	0.050	0.005	0.005	0.004	0.004	0.005	0.011	0.000	0.033	0.005
100	−0.8	0.831	0.832	0.244	0.247	0.220	0.220	0.224	0.227	0.042	0.039	0.035	0.037	0.039	0.120	0.068	0.113	0.096
	−0.7	0.680	0.681	0.146	0.145	0.131	0.130	0.140	0.141	0.038	0.036	0.032	0.034	0.037	0.087	0.035	0.071	0.066
	−0.6	0.566	0.569	0.103	0.103	0.105	0.105	0.100	0.100	0.052	0.050	0.046	0.046	0.050	0.084	0.029	0.063	0.070
	−0.5	0.435	0.436	0.090	0.090	0.086	0.085	0.086	0.087	0.060	0.057	0.054	0.050	0.057	0.079	0.031	0.059	0.074
	−0.4	0.327	0.328	0.101	0.099	0.077	0.077	0.091	0.093	0.059	0.054	0.055	0.049	0.055	0.068	0.029	0.053	0.069
	−0.3	0.210	0.213	0.099	0.099	0.076	0.077	0.090	0.090	0.058	0.054	0.053	0.047	0.053	0.063	0.025	0.047	0.063
	−0.2	0.135	0.133	0.103	0.102	0.072	0.071	0.086	0.088	0.059	0.054	0.055	0.046	0.054	0.063	0.025	0.052	0.063
	−0.1	0.071	0.071	0.074	0.073	0.059	0.060	0.075	0.078	0.059	0.056	0.054	0.048	0.055	0.066	0.022	0.048	0.061
	0	0.053	0.054	0.051	0.049	0.049	0.050	0.050	0.051	0.047	0.044	0.044	0.038	0.043	0.053	0.015	0.042	0.048
	0.1	0.036	0.035	0.038	0.037	0.046	0.046	0.046	0.047	0.039	0.036	0.038	0.031	0.036	0.042	0.010	0.037	0.039
	0.2	0.010	0.011	0.041	0.041	0.052	0.051	0.053	0.053	0.042	0.037	0.038	0.032	0.038	0.039	0.010	0.034	0.041
	0.3	0.021	0.021	0.049	0.050	0.056	0.055	0.059	0.058	0.058	0.054	0.053	0.047	0.053	0.053	0.020	0.034	0.058
	0.4	0.008	0.008	0.053	0.052	0.053	0.053	0.053	0.054	0.065	0.060	0.057	0.051	0.060	0.056	0.025	0.039	0.067
	0.5	0.004	0.004	0.055	0.055	0.053	0.052	0.057	0.059	0.054	0.050	0.048	0.044	0.051	0.045	0.025	0.038	0.059
	0.6	0.005	0.005	0.050	0.049	0.054	0.054	0.055	0.054	0.059	0.054	0.056	0.047	0.052	0.044	0.031	0.036	0.066
	0.7	0.006	0.006	0.052	0.053	0.049	0.048	0.055	0.057	0.060	0.055	0.054	0.049	0.054	0.041	0.031	0.031	0.064
	0.8	0.006	0.005	0.052	0.052	0.048	0.048	0.052	0.053	0.073	0.068	0.068	0.058	0.066	0.044	0.041	0.026	0.081

Table 3.3. Empirical size of the tests for model with an intercept only.

N	ϕ	BNM_0	$BEPO_0$	BNM_b	$BEPO_b$	BNM_a	$BEPO_a$	BNM_h	$BEPO_h$	MZ_a	MZ_t	MSB	P_t	MP_t	ADF	MZ_{aLS}	ADF_{LS}	MZ_{a2}
200	−0.8	0.908	0.908	0.097	0.098	0.101	0.101	0.104	0.102	0.023	0.021	0.019	0.021	0.022	0.085	0.021	0.070	0.042
	−0.7	0.768	0.767	0.076	0.078	0.073	0.076	0.083	0.082	0.039	0.038	0.033	0.036	0.038	0.075	0.023	0.059	0.050
	−0.6	0.598	0.599	0.065	0.066	0.067	0.068	0.073	0.072	0.049	0.047	0.042	0.044	0.047	0.069	0.027	0.053	0.057
	−0.5	0.481	0.480	0.057	0.057	0.061	0.062	0.065	0.062	0.056	0.055	0.049	0.051	0.055	0.064	0.029	0.051	0.061
	−0.4	0.346	0.346	0.055	0.057	0.055	0.057	0.066	0.065	0.054	0.050	0.048	0.047	0.051	0.057	0.030	0.046	0.059
	−0.3	0.202	0.202	0.061	0.061	0.052	0.052	0.058	0.055	0.060	0.056	0.055	0.051	0.055	0.059	0.034	0.051	0.063
	−0.2	0.151	0.151	0.071	0.071	0.054	0.056	0.067	0.066	0.052	0.049	0.046	0.045	0.048	0.052	0.028	0.049	0.055
	−0.1	0.085	0.086	0.070	0.074	0.057	0.057	0.066	0.065	0.053	0.050	0.050	0.045	0.049	0.053	0.028	0.047	0.055
	0	0.048	0.048	0.045	0.044	0.047	0.047	0.049	0.050	0.051	0.046	0.046	0.042	0.046	0.050	0.022	0.046	0.052
	0.1	0.033	0.033	0.038	0.040	0.048	0.047	0.046	0.044	0.045	0.041	0.040	0.037	0.041	0.043	0.015	0.037	0.044
	0.2	0.014	0.015	0.043	0.045	0.052	0.052	0.054	0.052	0.052	0.048	0.048	0.044	0.048	0.048	0.022	0.040	0.053
	0.3	0.005	0.005	0.048	0.048	0.047	0.046	0.052	0.052	0.058	0.054	0.051	0.047	0.053	0.052	0.031	0.047	0.059
	0.4	0.009	0.009	0.045	0.046	0.051	0.050	0.052	0.050	0.050	0.046	0.048	0.042	0.046	0.044	0.028	0.042	0.051
	0.5	0.006	0.006	0.045	0.046	0.046	0.047	0.052	0.053	0.050	0.046	0.048	0.042	0.046	0.041	0.027	0.038	0.053
	0.6	0.012	0.012	0.042	0.044	0.048	0.049	0.052	0.052	0.054	0.050	0.052	0.044	0.049	0.042	0.031	0.036	0.059
	0.7	0.009	0.009	0.042	0.043	0.046	0.046	0.047	0.047	0.059	0.053	0.054	0.048	0.052	0.041	0.038	0.037	0.064
	0.8	0.009	0.008	0.048	0.049	0.049	0.050	0.052	0.051	0.065	0.063	0.059	0.058	0.063	0.045	0.043	0.031	0.075
400	−0.8	0.936	0.947	0.074	0.077	0.073	0.070	0.072	0.073	0.028	0.027	0.022	0.026	0.027	0.085	0.012	0.066	0.035
	−0.7	0.781	0.793	0.063	0.064	0.062	0.061	0.064	0.066	0.046	0.044	0.038	0.041	0.044	0.066	0.021	0.061	0.050
	−0.6	0.625	0.633	0.060	0.064	0.055	0.055	0.059	0.059	0.052	0.049	0.044	0.047	0.048	0.061	0.033	0.059	0.056
	−0.5	0.456	0.480	0.054	0.055	0.058	0.056	0.062	0.061	0.052	0.049	0.047	0.047	0.049	0.054	0.033	0.052	0.054
	−0.4	0.275	0.285	0.060	0.061	0.054	0.053	0.057	0.055	0.054	0.051	0.050	0.048	0.050	0.051	0.037	0.052	0.055
	−0.3	0.179	0.182	0.054	0.055	0.050	0.049	0.054	0.054	0.051	0.049	0.048	0.047	0.049	0.050	0.031	0.049	0.052
	−0.2	0.096	0.115	0.056	0.057	0.051	0.049	0.057	0.055	0.056	0.052	0.049	0.048	0.051	0.053	0.034	0.050	0.057
	−0.1	0.066	0.086	0.070	0.069	0.053	0.053	0.063	0.063	0.051	0.047	0.050	0.045	0.048	0.049	0.032	0.049	0.052
	0	0.031	0.040	0.052	0.053	0.049	0.049	0.053	0.054	0.047	0.043	0.043	0.040	0.043	0.044	0.025	0.042	0.047
	0.1	0.016	0.022	0.041	0.042	0.049	0.049	0.046	0.048	0.053	0.050	0.050	0.047	0.049	0.051	0.029	0.046	0.053
	0.2	0.008	0.010	0.051	0.053	0.051	0.049	0.049	0.050	0.052	0.047	0.047	0.044	0.046	0.047	0.032	0.045	0.052
	0.3	0.012	0.012	0.053	0.053	0.055	0.053	0.055	0.056	0.052	0.048	0.049	0.045	0.047	0.047	0.033	0.045	0.054
	0.4	0.004	0.006	0.053	0.054	0.050	0.048	0.051	0.052	0.052	0.048	0.047	0.045	0.046	0.046	0.029	0.043	0.053
	0.5	0.003	0.008	0.051	0.053	0.047	0.047	0.053	0.053	0.051	0.048	0.045	0.045	0.047	0.044	0.032	0.042	0.053
	0.6	0.002	0.001	0.055	0.056	0.051	0.049	0.048	0.048	0.055	0.050	0.050	0.048	0.050	0.045	0.034	0.042	0.057
	0.7	0.002	0.002	0.049	0.052	0.049	0.047	0.045	0.046	0.053	0.051	0.050	0.047	0.050	0.043	0.040	0.042	0.058
	0.8	0.004	0.003	0.052	0.050	0.049	0.047	0.053	0.052	0.059	0.055	0.054	0.053	0.054	0.045	0.045	0.039	0.064

Table 3.4. Empirical size of the tests for model with an intercept and a trend.

N	ϕ	BNM_0	$BEPO_0$	BNM_b	$BEPO_b$	BNM_a	$BEPO_a$	BNM_h	$BEPO_h$	MZ_a	MZ_t	MSB	P_t	MP_t	ADF	MZ_{aLS}	ADF_{LS}	MZ_{a2}
50	−0.8	0.778	0.781	0.727	0.726	0.532	0.533	0.643	0.644	0.993	0.991	0.994	0.957	0.988	1.000	0.910	1.000	0.957
	−0.7	0.717	0.717	0.620	0.618	0.437	0.438	0.534	0.535	0.931	0.924	0.933	0.864	0.917	0.995	0.716	0.989	0.876
	−0.6	0.613	0.613	0.491	0.489	0.329	0.329	0.415	0.416	0.761	0.749	0.768	0.693	0.745	0.959	0.453	0.922	0.699
	−0.5	0.456	0.457	0.375	0.372	0.237	0.237	0.307	0.307	0.523	0.514	0.531	0.473	0.511	0.840	0.234	0.771	0.469
	−0.4	0.337	0.341	0.267	0.264	0.175	0.176	0.215	0.216	0.306	0.300	0.315	0.281	0.302	0.671	0.096	0.579	0.270
	−0.3	0.224	0.225	0.180	0.178	0.119	0.119	0.143	0.143	0.160	0.155	0.167	0.148	0.157	0.468	0.032	0.382	0.134
	−0.2	0.139	0.139	0.112	0.110	0.072	0.072	0.096	0.097	0.071	0.069	0.075	0.068	0.071	0.298	0.012	0.233	0.058
	−0.1	0.069	0.071	0.069	0.069	0.067	0.067	0.066	0.066	0.028	0.027	0.031	0.028	0.028	0.176	0.002	0.132	0.023
	0	0.043	0.044	0.048	0.048	0.053	0.053	0.057	0.056	0.010	0.010	0.010	0.011	0.011	0.090	0.001	0.068	0.009
	0.1	0.025	0.026	0.042	0.041	0.053	0.052	0.050	0.051	0.005	0.005	0.005	0.006	0.005	0.056	0.000	0.045	0.004
	0.2	0.015	0.015	0.040	0.038	0.060	0.059	0.055	0.055	0.001	0.001	0.002	0.002	0.002	0.030	0.000	0.028	0.001
	0.3	0.003	0.003	0.054	0.053	0.058	0.057	0.056	0.056	0.001	0.001	0.001	0.001	0.001	0.018	0.000	0.020	0.001
	0.4	0.006	0.006	0.051	0.050	0.051	0.051	0.060	0.060	0.000	0.000	0.000	0.000	0.000	0.011	0.000	0.016	0.000
	0.5	0.007	0.007	0.057	0.055	0.049	0.049	0.056	0.056	0.000	0.000	0.000	0.000	0.000	0.007	0.000	0.012	0.000
	0.6	0.002	0.002	0.049	0.049	0.040	0.038	0.054	0.054	0.000	0.000	0.000	0.000	0.000	0.005	0.000	0.013	0.000
	0.7	0.002	0.002	0.050	0.050	0.047	0.047	0.054	0.054	0.000	0.000	0.000	0.000	0.000	0.004	0.000	0.008	0.000
	0.8	0.001	0.001	0.049	0.048	0.039	0.038	0.049	0.049	0.000	0.000	0.000	0.000	0.000	0.004	0.000	0.011	0.000
100	−0.8	0.931	0.931	0.595	0.596	0.465	0.463	0.509	0.509	0.067	0.068	0.067	0.070	0.069	0.127	0.143	0.196	0.163
	−0.7	0.841	0.838	0.343	0.344	0.262	0.260	0.281	0.281	0.040	0.040	0.040	0.042	0.042	0.089	0.073	0.116	0.091
	−0.6	0.743	0.742	0.211	0.212	0.180	0.178	0.180	0.180	0.036	0.037	0.036	0.037	0.037	0.076	0.045	0.078	0.064
	−0.5	0.581	0.579	0.185	0.186	0.140	0.138	0.146	0.144	0.035	0.035	0.035	0.035	0.036	0.070	0.036	0.065	0.052
	−0.4	0.380	0.378	0.177	0.177	0.111	0.110	0.135	0.134	0.038	0.038	0.039	0.039	0.040	0.066	0.032	0.061	0.051
	−0.3	0.272	0.269	0.161	0.162	0.103	0.102	0.115	0.115	0.034	0.034	0.035	0.034	0.035	0.059	0.027	0.054	0.043
	−0.2	0.172	0.167	0.128	0.129	0.085	0.084	0.098	0.098	0.033	0.033	0.034	0.032	0.033	0.057	0.023	0.052	0.039
	−0.1	0.089	0.088	0.080	0.080	0.062	0.061	0.068	0.067	0.028	0.028	0.028	0.026	0.029	0.054	0.016	0.054	0.031
	0	0.050	0.050	0.046	0.047	0.053	0.052	0.045	0.044	0.020	0.021	0.020	0.021	0.022	0.043	0.010	0.042	0.021
	0.1	0.028	0.027	0.041	0.041	0.048	0.047	0.044	0.043	0.011	0.011	0.012	0.011	0.012	0.025	0.005	0.029	0.010
	0.2	0.010	0.010	0.043	0.044	0.055	0.054	0.050	0.049	0.012	0.012	0.013	0.011	0.012	0.019	0.004	0.018	0.012
	0.3	0.005	0.005	0.061	0.062	0.056	0.055	0.050	0.050	0.020	0.020	0.022	0.020	0.021	0.023	0.011	0.022	0.020
	0.4	0.006	0.006	0.060	0.061	0.055	0.054	0.047	0.047	0.031	0.029	0.033	0.028	0.030	0.026	0.020	0.023	0.034
	0.5	0.004	0.004	0.050	0.051	0.049	0.048	0.046	0.045	0.041	0.040	0.042	0.039	0.040	0.032	0.028	0.026	0.045
	0.6	0.006	0.006	0.056	0.056	0.051	0.051	0.048	0.048	0.040	0.040	0.042	0.038	0.040	0.027	0.034	0.023	0.050
	0.7	0.002	0.002	0.054	0.053	0.050	0.049	0.044	0.044	0.043	0.041	0.044	0.041	0.042	0.021	0.035	0.019	0.049
	0.8	0.001	0.000	0.047	0.047	0.045	0.044	0.047	0.047	0.056	0.054	0.059	0.052	0.054	0.018	0.048	0.015	0.067

Table 3.5. Empirical size of the tests for model with an intercept and a trend.

N	ϕ	BNM_0	$BEPO_0$	BNM_b	$BEPO_b$	BNM_a	$BEPO_a$	BNM_h	$BEPO_h$	MZ_a	MZ_t	MSB	P_t	MP_t	ADF	MZ_{aLS}	ADF_{LS}	MZ_{a2}
200	−0.8	0.983	0.983	0.187	0.187	0.190	0.191	0.189	0.188	0.015	0.015	0.015	0.016	0.016	0.074	0.042	0.095	0.049
	−0.7	0.921	0.921	0.112	0.112	0.113	0.114	0.122	0.120	0.021	0.021	0.021	0.022	0.022	0.061	0.032	0.071	0.042
	−0.6	0.796	0.796	0.095	0.095	0.088	0.088	0.098	0.096	0.026	0.026	0.026	0.026	0.027	0.052	0.031	0.058	0.039
	−0.5	0.625	0.624	0.076	0.076	0.076	0.077	0.084	0.082	0.032	0.033	0.032	0.034	0.034	0.051	0.035	0.054	0.043
	−0.4	0.458	0.457	0.069	0.069	0.067	0.068	0.073	0.072	0.031	0.030	0.032	0.031	0.032	0.043	0.030	0.047	0.039
	−0.3	0.284	0.283	0.084	0.084	0.068	0.069	0.077	0.075	0.040	0.040	0.040	0.039	0.041	0.048	0.034	0.050	0.045
	−0.2	0.161	0.161	0.098	0.098	0.069	0.069	0.083	0.082	0.035	0.035	0.036	0.035	0.037	0.046	0.031	0.046	0.040
	−0.1	0.086	0.085	0.084	0.084	0.059	0.059	0.076	0.075	0.034	0.033	0.034	0.032	0.034	0.042	0.026	0.043	0.036
	0	0.055	0.055	0.046	0.046	0.047	0.047	0.053	0.052	0.027	0.027	0.028	0.027	0.028	0.039	0.020	0.042	0.028
	0.1	0.026	0.026	0.036	0.036	0.049	0.049	0.046	0.046	0.017	0.017	0.018	0.016	0.017	0.022	0.011	0.026	0.016
	0.2	0.013	0.013	0.051	0.051	0.047	0.048	0.053	0.051	0.025	0.025	0.026	0.024	0.025	0.028	0.016	0.027	0.025
	0.3	0.011	0.011	0.050	0.050	0.054	0.054	0.053	0.052	0.037	0.035	0.038	0.034	0.035	0.035	0.030	0.036	0.038
	0.4	0.003	0.003	0.052	0.052	0.050	0.050	0.054	0.053	0.037	0.036	0.038	0.035	0.036	0.033	0.033	0.033	0.042
	0.5	0.004	0.004	0.049	0.049	0.046	0.047	0.055	0.054	0.032	0.032	0.034	0.031	0.032	0.025	0.027	0.027	0.036
	0.6	0.002	0.002	0.046	0.046	0.048	0.049	0.053	0.052	0.039	0.038	0.040	0.037	0.038	0.027	0.033	0.024	0.043
	0.7	0.000	0.000	0.049	0.049	0.047	0.048	0.051	0.050	0.046	0.046	0.048	0.044	0.046	0.023	0.040	0.024	0.050
	0.8	0.004	0.004	0.048	0.048	0.049	0.050	0.052	0.051	0.055	0.054	0.056	0.052	0.053	0.018	0.054	0.019	0.065
400	−0.8	0.999	0.999	0.101	0.102	0.104	0.102	0.103	0.103	0.010	0.011	0.010	0.011	0.011	0.062	0.014	0.071	0.019
	−0.7	0.956	0.956	0.078	0.078	0.076	0.075	0.086	0.085	0.017	0.017	0.017	0.019	0.019	0.053	0.020	0.054	0.024
	−0.6	0.839	0.839	0.062	0.062	0.071	0.070	0.073	0.072	0.031	0.032	0.031	0.032	0.032	0.051	0.031	0.055	0.037
	−0.5	0.650	0.650	0.061	0.061	0.063	0.061	0.064	0.063	0.034	0.035	0.036	0.035	0.035	0.045	0.032	0.048	0.040
	−0.4	0.487	0.487	0.056	0.056	0.061	0.060	0.056	0.056	0.039	0.040	0.040	0.040	0.041	0.047	0.035	0.048	0.044
	−0.3	0.307	0.309	0.052	0.052	0.059	0.058	0.058	0.058	0.041	0.042	0.041	0.041	0.043	0.045	0.034	0.048	0.043
	−0.2	0.205	0.205	0.067	0.067	0.055	0.054	0.056	0.056	0.038	0.037	0.037	0.037	0.038	0.040	0.031	0.041	0.039
	−0.1	0.115	0.115	0.081	0.081	0.060	0.059	0.069	0.068	0.035	0.036	0.036	0.036	0.036	0.039	0.030	0.040	0.037
	0	0.055	0.056	0.048	0.049	0.046	0.046	0.051	0.050	0.036	0.035	0.037	0.034	0.036	0.040	0.028	0.043	0.036
	0.1	0.039	0.039	0.038	0.038	0.050	0.050	0.051	0.050	0.028	0.028	0.028	0.028	0.029	0.029	0.019	0.033	0.027
	0.2	0.015	0.015	0.055	0.055	0.051	0.051	0.051	0.050	0.039	0.039	0.040	0.038	0.039	0.039	0.036	0.043	0.041
	0.3	0.011	0.011	0.047	0.047	0.052	0.051	0.054	0.054	0.037	0.037	0.039	0.037	0.038	0.035	0.032	0.037	0.039
	0.4	0.007	0.007	0.045	0.045	0.047	0.046	0.055	0.055	0.035	0.034	0.035	0.033	0.034	0.031	0.034	0.035	0.036
	0.5	0.005	0.005	0.046	0.046	0.052	0.050	0.054	0.053	0.038	0.037	0.040	0.038	0.039	0.030	0.034	0.033	0.041
	0.6	0.002	0.002	0.045	0.045	0.051	0.050	0.051	0.050	0.041	0.040	0.041	0.041	0.042	0.032	0.040	0.035	0.044
	0.7	0.003	0.003	0.050	0.050	0.046	0.045	0.054	0.053	0.045	0.045	0.046	0.044	0.046	0.030	0.046	0.031	0.051
	0.8	0.003	0.003	0.044	0.044	0.046	0.046	0.052	0.052	0.053	0.052	0.055	0.051	0.053	0.030	0.059	0.031	0.061

Table 3.6. Power of the tests for model with an intercept only, $\phi = 0$.

N	ρ	BNM_0	$BEPO_0$	BNM_b	$BEPO_b$	BNM_a	$BEPO_a$	BNM_h	$BEPO_h$	MZ_a	MZ_t	MSB	P_t	MP_t	ADF	MZ_{aLS}	ADF_{LS}	MZ_{a2}
50	0.80	0.550	0.566	0.530	0.542	0.336	0.338	0.408	0.424	0.772	0.753	0.713	0.737	0.748	0.862	0.216	0.383	0.769
	0.82	0.486	0.495	0.484	0.488	0.312	0.313	0.374	0.388	0.702	0.684	0.640	0.657	0.680	0.812	0.178	0.324	0.696
	0.84	0.430	0.440	0.419	0.425	0.268	0.268	0.312	0.326	0.614	0.591	0.548	0.564	0.589	0.741	0.139	0.266	0.605
	0.86	0.350	0.358	0.346	0.355	0.239	0.237	0.264	0.275	0.536	0.515	0.471	0.481	0.513	0.666	0.110	0.221	0.526
	0.88	0.282	0.292	0.298	0.299	0.207	0.206	0.234	0.243	0.448	0.434	0.383	0.395	0.430	0.580	0.083	0.181	0.435
	0.90	0.211	0.222	0.241	0.244	0.179	0.179	0.188	0.196	0.366	0.346	0.319	0.309	0.345	0.488	0.061	0.141	0.355
	0.92	0.184	0.187	0.195	0.195	0.155	0.155	0.157	0.163	0.283	0.269	0.249	0.229	0.269	0.392	0.046	0.121	0.271
	0.94	0.125	0.128	0.149	0.148	0.120	0.120	0.121	0.126	0.213	0.198	0.182	0.164	0.198	0.304	0.033	0.106	0.200
	0.96	0.100	0.103	0.112	0.112	0.101	0.100	0.093	0.093	0.142	0.131	0.120	0.105	0.131	0.215	0.020	0.084	0.131
	0.98	0.064	0.068	0.079	0.078	0.074	0.073	0.066	0.066	0.094	0.086	0.086	0.068	0.085	0.146	0.013	0.067	0.087
100	0.80	0.926	0.930	0.913	0.922	0.874	0.882	0.901	0.911	0.841	0.836	0.823	0.831	0.834	0.855	0.646	0.597	0.872
	0.82	0.884	0.887	0.889	0.897	0.839	0.851	0.881	0.892	0.831	0.824	0.809	0.816	0.820	0.844	0.577	0.532	0.859
	0.84	0.851	0.853	0.846	0.858	0.797	0.811	0.848	0.866	0.816	0.805	0.786	0.795	0.801	0.827	0.501	0.460	0.843
	0.86	0.819	0.823	0.797	0.813	0.736	0.750	0.790	0.808	0.783	0.769	0.742	0.757	0.765	0.798	0.404	0.367	0.806
	0.88	0.739	0.747	0.715	0.733	0.657	0.672	0.711	0.733	0.717	0.702	0.668	0.684	0.699	0.743	0.315	0.290	0.742
	0.90	0.626	0.631	0.607	0.625	0.553	0.567	0.609	0.632	0.632	0.614	0.572	0.588	0.610	0.660	0.236	0.225	0.651
	0.92	0.514	0.519	0.472	0.491	0.439	0.453	0.478	0.503	0.487	0.472	0.431	0.443	0.468	0.521	0.156	0.145	0.503
	0.94	0.328	0.336	0.332	0.350	0.313	0.322	0.337	0.358	0.353	0.337	0.303	0.311	0.335	0.388	0.100	0.103	0.362
	0.96	0.237	0.240	0.202	0.208	0.205	0.210	0.222	0.231	0.211	0.202	0.181	0.180	0.201	0.242	0.060	0.070	0.217
	0.98	0.097	0.098	0.103	0.107	0.114	0.117	0.118	0.121	0.110	0.104	0.093	0.090	0.103	0.124	0.033	0.052	0.112
200	0.80	0.999	0.999	0.997	0.997	0.986	0.987	0.996	0.996	0.952	0.949	0.944	0.947	0.948	0.970	0.887	0.842	0.964
	0.82	0.996	0.996	0.995	0.995	0.985	0.986	0.995	0.996	0.954	0.950	0.944	0.949	0.949	0.969	0.881	0.832	0.964
	0.84	0.996	0.996	0.994	0.996	0.984	0.985	0.992	0.994	0.951	0.949	0.942	0.946	0.947	0.963	0.876	0.823	0.964
	0.86	0.994	0.994	0.988	0.990	0.980	0.983	0.988	0.989	0.947	0.943	0.937	0.940	0.943	0.956	0.859	0.790	0.957
	0.88	0.988	0.988	0.983	0.986	0.970	0.974	0.978	0.981	0.941	0.937	0.931	0.933	0.934	0.947	0.828	0.741	0.953
	0.90	0.970	0.971	0.963	0.968	0.943	0.950	0.966	0.970	0.930	0.924	0.916	0.919	0.922	0.929	0.758	0.647	0.944
	0.92	0.903	0.905	0.912	0.923	0.886	0.899	0.916	0.922	0.897	0.888	0.876	0.877	0.882	0.895	0.613	0.496	0.908
	0.94	0.810	0.811	0.767	0.791	0.735	0.757	0.777	0.792	0.793	0.778	0.745	0.755	0.772	0.791	0.410	0.310	0.805
	0.96	0.545	0.548	0.502	0.524	0.475	0.500	0.521	0.540	0.523	0.507	0.465	0.480	0.504	0.529	0.209	0.154	0.529
	0.98	0.223	0.224	0.206	0.213	0.207	0.217	0.224	0.229	0.220	0.208	0.190	0.192	0.207	0.221	0.085	0.071	0.219
400	0.80	1.000	1.000	1.000	1.000	0.999	0.999	1.000	1.000	1.000	0.999	0.999	0.999	0.999	1.000	0.991	0.987	1.000
	0.82	1.000	1.000	1.000	1.000	0.999	0.999	1.000	1.000	1.000	0.999	0.999	0.999	0.999	1.000	0.990	0.983	1.000
	0.84	1.000	1.000	1.000	1.000	0.999	0.999	1.000	1.000	1.000	0.999	0.999	0.999	0.999	1.000	0.986	0.975	1.000
	0.86	1.000	1.000	1.000	1.000	0.999	0.999	1.000	1.000	0.999	0.998	0.998	0.997	0.997	0.999	0.985	0.970	0.999
	0.88	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.998	0.997	0.998	0.997	0.997	0.999	0.980	0.956	0.999
	0.90	0.999	0.999	1.000	1.000	0.999	0.999	1.000	1.000	0.997	0.996	0.997	0.995	0.995	0.998	0.974	0.945	0.998
	0.92	0.999	1.000	0.999	1.000	0.999	0.999	0.999	0.999	0.995	0.993	0.993	0.992	0.992	0.995	0.963	0.918	0.996
	0.94	0.994	0.996	0.997	0.998	0.993	0.993	0.996	0.997	0.988	0.985	0.984	0.982	0.984	0.985	0.915	0.829	0.990
	0.96	0.889	0.936	0.950	0.962	0.936	0.942	0.949	0.958	0.949	0.938	0.931	0.931	0.935	0.940	0.700	0.547	0.952
	0.98	0.415	0.489	0.544	0.575	0.516	0.529	0.543	0.568	0.559	0.542	0.507	0.527	0.539	0.551	0.258	0.170	0.560

Table 3.7. Power of the tests for model with an intercept and a trend, $\phi = 0$.

N	ρ	BNM_0	$BEPO_0$	BNM_b	$BEPO_b$	BNM_a	$BEPO_a$	BNM_h	$BEPO_h$	MZ_a	MZ_t	MSB	P_t	MP_t	ADF	MZ_{aLS}	ADF_{LS}	MZ_{a2}
50	0.80	0.272	0.275	0.236	0.234	0.173	0.173	0.213	0.213	0.092	0.095	0.095	0.106	0.101	0.468	0.009	0.262	0.082
	0.82	0.217	0.218	0.198	0.196	0.159	0.159	0.186	0.186	0.075	0.077	0.074	0.089	0.082	0.416	0.007	0.238	0.066
	0.84	0.210	0.210	0.170	0.167	0.140	0.141	0.161	0.161	0.057	0.057	0.059	0.064	0.059	0.349	0.005	0.196	0.051
	0.86	0.151	0.153	0.139	0.136	0.120	0.119	0.134	0.134	0.046	0.047	0.048	0.054	0.050	0.300	0.005	0.171	0.041
	0.88	0.115	0.116	0.111	0.110	0.110	0.109	0.112	0.113	0.037	0.038	0.036	0.042	0.040	0.250	0.003	0.144	0.031
	0.90	0.092	0.092	0.093	0.091	0.090	0.090	0.094	0.093	0.029	0.029	0.030	0.032	0.031	0.209	0.002	0.125	0.025
	0.92	0.092	0.093	0.082	0.081	0.080	0.080	0.087	0.086	0.020	0.020	0.022	0.024	0.022	0.172	0.002	0.106	0.019
	0.94	0.064	0.064	0.069	0.069	0.070	0.069	0.069	0.069	0.017	0.017	0.018	0.019	0.017	0.144	0.002	0.097	0.014
	0.96	0.045	0.045	0.050	0.049	0.067	0.066	0.066	0.066	0.012	0.013	0.013	0.015	0.014	0.117	0.001	0.082	0.010
	0.98	0.052	0.053	0.043	0.042	0.056	0.056	0.056	0.056	0.010	0.010	0.010	0.011	0.011	0.094	0.001	0.071	0.009
100	0.80	0.684	0.683	0.703	0.705	0.572	0.569	0.628	0.627	0.505	0.509	0.498	0.519	0.515	0.632	0.339	0.457	0.546
	0.82	0.623	0.615	0.633	0.635	0.508	0.506	0.554	0.553	0.429	0.433	0.423	0.438	0.437	0.573	0.262	0.392	0.458
	0.84	0.570	0.564	0.543	0.547	0.434	0.432	0.481	0.478	0.343	0.344	0.342	0.349	0.347	0.500	0.195	0.312	0.368
	0.86	0.478	0.471	0.458	0.461	0.364	0.361	0.395	0.393	0.260	0.261	0.259	0.265	0.266	0.409	0.140	0.245	0.279
	0.88	0.356	0.355	0.368	0.370	0.297	0.294	0.315	0.315	0.186	0.186	0.184	0.189	0.191	0.321	0.094	0.185	0.201
	0.90	0.281	0.273	0.278	0.281	0.220	0.218	0.230	0.228	0.133	0.135	0.132	0.137	0.139	0.250	0.069	0.146	0.145
	0.92	0.198	0.192	0.200	0.201	0.166	0.163	0.168	0.166	0.088	0.088	0.088	0.088	0.090	0.168	0.044	0.102	0.094
	0.94	0.123	0.121	0.131	0.132	0.122	0.120	0.111	0.110	0.054	0.055	0.054	0.054	0.056	0.115	0.027	0.077	0.057
	0.96	0.092	0.091	0.091	0.092	0.086	0.086	0.077	0.076	0.037	0.037	0.037	0.038	0.039	0.081	0.016	0.056	0.039
	0.98	0.074	0.072	0.062	0.062	0.065	0.064	0.051	0.051	0.023	0.024	0.025	0.024	0.025	0.052	0.013	0.046	0.025
200	0.80	0.964	0.964	0.968	0.968	0.947	0.948	0.967	0.965	0.844	0.844	0.844	0.845	0.845	0.845	0.844	0.800	0.890
	0.82	0.960	0.959	0.953	0.953	0.929	0.930	0.952	0.951	0.839	0.838	0.837	0.838	0.838	0.836	0.818	0.769	0.879
	0.84	0.926	0.926	0.933	0.933	0.902	0.903	0.928	0.926	0.829	0.828	0.829	0.830	0.829	0.829	0.789	0.748	0.866
	0.86	0.885	0.883	0.887	0.887	0.855	0.856	0.894	0.892	0.792	0.792	0.792	0.790	0.790	0.796	0.710	0.672	0.828
	0.88	0.849	0.848	0.824	0.824	0.784	0.786	0.827	0.823	0.724	0.723	0.721	0.723	0.725	0.747	0.609	0.583	0.757
	0.90	0.718	0.718	0.722	0.722	0.660	0.662	0.710	0.705	0.603	0.603	0.601	0.603	0.608	0.654	0.470	0.457	0.634
	0.92	0.566	0.564	0.570	0.570	0.514	0.516	0.567	0.562	0.431	0.431	0.432	0.431	0.436	0.501	0.306	0.312	0.452
	0.94	0.374	0.371	0.370	0.370	0.331	0.333	0.382	0.376	0.253	0.256	0.253	0.253	0.261	0.318	0.171	0.191	0.263
	0.96	0.183	0.182	0.193	0.193	0.185	0.186	0.198	0.195	0.119	0.120	0.117	0.119	0.123	0.158	0.078	0.099	0.124
	0.98	0.070	0.069	0.086	0.086	0.084	0.086	0.093	0.091	0.049	0.049	0.049	0.049	0.049	0.068	0.035	0.055	0.052
400	0.80	1.000	1.000	0.999	0.999	0.997	0.997	1.000	1.000	0.974	0.973	0.975	0.973	0.973	0.984	0.965	0.943	0.986
	0.82	1.000	1.000	0.999	0.999	0.997	0.997	0.998	0.998	0.972	0.971	0.972	0.970	0.970	0.982	0.964	0.936	0.984
	0.84	0.999	0.999	0.998	0.998	0.996	0.996	0.998	0.998	0.966	0.965	0.966	0.966	0.965	0.971	0.957	0.924	0.978
	0.86	0.996	0.996	0.996	0.996	0.991	0.991	0.995	0.995	0.963	0.963	0.964	0.963	0.963	0.966	0.953	0.917	0.975
	0.88	0.992	0.992	0.991	0.991	0.985	0.985	0.988	0.988	0.956	0.955	0.956	0.955	0.955	0.954	0.945	0.901	0.969
	0.90	0.980	0.980	0.977	0.977	0.971	0.971	0.975	0.975	0.948	0.947	0.948	0.947	0.947	0.942	0.931	0.883	0.963
	0.92	0.955	0.956	0.941	0.941	0.933	0.933	0.943	0.943	0.921	0.923	0.921	0.923	0.925	0.916	0.884	0.819	0.938
	0.94	0.851	0.851	0.844	0.845	0.825	0.824	0.849	0.847	0.816	0.817	0.813	0.816	0.819	0.822	0.709	0.633	0.834
	0.96	0.643	0.644	0.580	0.580	0.568	0.567	0.584	0.581	0.500	0.502	0.499	0.502	0.506	0.532	0.377	0.325	0.511
	0.98	0.228	0.228	0.199	0.199	0.191	0.191	0.211	0.208	0.145	0.146	0.146	0.146	0.148	0.163	0.107	0.102	0.147

Table 3.8. Size-adjusted power of the tests for model with an intercept only, $\phi = -0.5$.

N	ρ	BNM_0	$BEPO_0$	BNM_b	$BEPO_b$	BNM_a	$BEPO_a$	BNM_h	$BEPO_h$	MZ_a	MZ_t	MSB	P_t	MP_t	ADF	MZ_{aLS}	ADF_{LS}	MZ_{a2}
50	0.80	0.389	0.394	0.411	0.417	0.307	0.308	0.397	0.392	0.657	0.670	0.606	0.689	0.605	0.652	0.555	0.430	0.707
	0.82	0.399	0.406	0.387	0.394	0.242	0.246	0.355	0.350	0.585	0.602	0.533	0.625	0.545	0.582	0.487	0.373	0.633
	0.84	0.34	0.348	0.335	0.344	0.234	0.234	0.332	0.327	0.496	0.514	0.448	0.553	0.473	0.493	0.408	0.312	0.539
	0.86	0.256	0.269	0.290	0.295	0.226	0.226	0.282	0.278	0.428	0.447	0.383	0.474	0.415	0.425	0.353	0.263	0.467
	0.88	0.221	0.226	0.257	0.262	0.179	0.179	0.262	0.260	0.353	0.371	0.309	0.399	0.347	0.352	0.291	0.216	0.379
	0.90	0.201	0.205	0.210	0.215	0.191	0.193	0.192	0.188	0.284	0.299	0.256	0.316	0.282	0.282	0.235	0.174	0.307
	0.92	0.139	0.144	0.157	0.163	0.141	0.143	0.168	0.164	0.224	0.234	0.197	0.245	0.219	0.223	0.195	0.144	0.231
	0.94	0.118	0.119	0.138	0.142	0.120	0.120	0.130	0.125	0.166	0.174	0.150	0.174	0.162	0.166	0.149	0.119	0.170
	0.96	0.093	0.1	0.110	0.114	0.101	0.099	0.087	0.086	0.110	0.114	0.103	0.115	0.108	0.110	0.105	0.092	0.111
	0.98	0.055	0.056	0.084	0.086	0.071	0.071	0.078	0.075	0.073	0.076	0.068	0.074	0.071	0.072	0.067	0.064	0.074
100	0.80	0.726	0.725	0.810	0.813	0.776	0.784	0.801	0.806	0.612	0.616	0.603	0.633	0.615	0.710	0.539	0.485	0.640
	0.82	0.695	0.693	0.768	0.774	0.741	0.749	0.754	0.762	0.598	0.599	0.591	0.615	0.598	0.690	0.512	0.443	0.625
	0.84	0.697	0.695	0.716	0.723	0.689	0.699	0.711	0.719	0.574	0.577	0.562	0.594	0.577	0.666	0.476	0.399	0.597
	0.86	0.64	0.637	0.643	0.649	0.639	0.647	0.647	0.654	0.551	0.553	0.536	0.569	0.551	0.626	0.429	0.344	0.566
	0.88	0.583	0.579	0.567	0.573	0.555	0.565	0.571	0.580	0.513	0.518	0.495	0.534	0.516	0.574	0.386	0.287	0.521
	0.90	0.508	0.501	0.483	0.490	0.475	0.483	0.487	0.494	0.444	0.450	0.435	0.466	0.450	0.500	0.333	0.239	0.459
	0.92	0.41	0.405	0.379	0.383	0.382	0.389	0.382	0.389	0.363	0.367	0.353	0.382	0.368	0.402	0.260	0.178	0.372
	0.94	0.299	0.297	0.281	0.283	0.278	0.286	0.277	0.283	0.281	0.285	0.270	0.297	0.286	0.309	0.201	0.127	0.285
	0.96	0.192	0.191	0.177	0.180	0.184	0.188	0.181	0.186	0.189	0.192	0.182	0.198	0.193	0.202	0.140	0.092	0.191
	0.98	0.115	0.111	0.108	0.109	0.104	0.105	0.101	0.104	0.109	0.110	0.107	0.111	0.112	0.113	0.099	0.070	0.112
200	0.80	0.914	0.914	0.985	0.986	0.968	0.971	0.978	0.978	0.864	0.864	0.855	0.873	0.864	0.969	0.792	0.772	0.884
	0.82	0.911	0.911	0.982	0.984	0.961	0.965	0.976	0.976	0.865	0.866	0.853	0.875	0.867	0.963	0.779	0.752	0.882
	0.84	0.897	0.897	0.979	0.981	0.957	0.961	0.971	0.972	0.856	0.857	0.844	0.866	0.858	0.947	0.773	0.726	0.875
	0.86	0.886	0.886	0.973	0.976	0.949	0.956	0.966	0.966	0.845	0.844	0.837	0.851	0.843	0.929	0.747	0.682	0.863
	0.88	0.872	0.873	0.955	0.959	0.927	0.934	0.947	0.951	0.836	0.834	0.825	0.840	0.833	0.909	0.721	0.628	0.852
	0.90	0.831	0.832	0.933	0.940	0.891	0.901	0.915	0.917	0.812	0.810	0.799	0.819	0.810	0.882	0.673	0.555	0.825
	0.92	0.788	0.789	0.862	0.875	0.823	0.838	0.852	0.857	0.771	0.769	0.753	0.780	0.769	0.828	0.603	0.446	0.778
	0.94	0.677	0.68	0.727	0.744	0.690	0.711	0.707	0.716	0.675	0.677	0.653	0.687	0.678	0.723	0.473	0.306	0.677
	0.96	0.501	0.504	0.489	0.511	0.460	0.483	0.476	0.483	0.467	0.468	0.441	0.477	0.472	0.499	0.306	0.172	0.468
	0.98	0.237	0.239	0.217	0.225	0.205	0.216	0.203	0.206	0.217	0.218	0.208	0.221	0.219	0.227	0.160	0.089	0.217
400	0.80	0.99	0.99	0.999	1.000	0.994	0.994	0.998	0.998	0.994	0.993	0.993	0.992	0.992	1.000	0.974	0.992	0.995
	0.82	0.989	0.989	1.000	1.000	0.996	0.996	0.999	0.999	0.994	0.994	0.994	0.994	0.993	1.000	0.974	0.988	0.996
	0.84	0.987	0.987	1.000	1.000	0.996	0.996	0.999	0.999	0.993	0.992	0.993	0.992	0.992	1.000	0.971	0.981	0.995
	0.86	0.987	0.987	1.000	1.000	0.996	0.997	0.999	0.999	0.993	0.992	0.992	0.991	0.991	1.000	0.971	0.973	0.995
	0.88	0.982	0.981	1.000	1.000	0.996	0.997	0.999	0.999	0.990	0.988	0.989	0.988	0.987	0.998	0.961	0.950	0.991
	0.90	0.978	0.981	0.999	0.999	0.996	0.996	0.999	0.999	0.989	0.986	0.989	0.985	0.985	0.998	0.955	0.926	0.991
	0.92	0.962	0.963	0.998	0.998	0.994	0.995	0.996	0.996	0.981	0.979	0.980	0.976	0.976	0.992	0.934	0.871	0.984
	0.94	0.942	0.942	0.990	0.990	0.984	0.988	0.989	0.990	0.963	0.959	0.961	0.957	0.955	0.976	0.882	0.761	0.968
	0.96	0.851	0.854	0.925	0.933	0.910	0.921	0.922	0.932	0.894	0.890	0.885	0.886	0.885	0.906	0.727	0.514	0.900
	0.98	0.504	0.502	0.522	0.533	0.507	0.528	0.507	0.527	0.519	0.519	0.497	0.519	0.521	0.527	0.354	0.186	0.525

Table 3.9. Size-adjusted power of the tests for model with an intercept and a trend, $\phi = -0.5$.

N	ρ	BNM_0	$BEPO_0$	BNM_b	$BEPO_b$	BNM_a	$BEPO_a$	BNM_h	$BEPO_h$	MZ_a	MZ_t	MSB	P_t	MP_t	ADF	MZ_{aLS}	ADF_{LS}	MZ_{a2}
50	0.80	0.211	0.209	0.211	0.212	0.151	0.151	0.159	0.172	0.285	0.277	0.265	0.293	0.266	0.274	0.291	0.207	0.313
	0.82	0.196	0.193	0.189	0.192	0.128	0.128	0.155	0.165	0.239	0.235	0.227	0.248	0.230	0.232	0.245	0.181	0.262
	0.84	0.169	0.166	0.164	0.165	0.113	0.114	0.130	0.139	0.204	0.203	0.190	0.207	0.193	0.196	0.206	0.156	0.221
	0.86	0.124	0.120	0.134	0.135	0.105	0.104	0.114	0.122	0.173	0.171	0.160	0.173	0.166	0.170	0.173	0.135	0.179
	0.88	0.097	0.097	0.116	0.117	0.083	0.084	0.099	0.105	0.137	0.136	0.130	0.139	0.135	0.134	0.139	0.110	0.143
	0.90	0.081	0.076	0.097	0.098	0.075	0.076	0.081	0.087	0.112	0.112	0.105	0.110	0.113	0.110	0.110	0.098	0.112
	0.92	0.074	0.073	0.084	0.086	0.070	0.070	0.073	0.078	0.089	0.089	0.088	0.088	0.088	0.091	0.090	0.081	0.090
	0.94	0.074	0.074	0.071	0.072	0.063	0.063	0.055	0.060	0.073	0.072	0.073	0.070	0.072	0.072	0.074	0.067	0.074
	0.96	0.062	0.061	0.059	0.059	0.054	0.054	0.052	0.056	0.061	0.059	0.060	0.059	0.059	0.060	0.061	0.058	0.060
	0.98	0.053	0.051	0.053	0.053	0.051	0.051	0.047	0.050	0.049	0.048	0.048	0.045	0.048	0.049	0.048	0.050	0.050
100	0.80	0.603	0.585	0.583	0.586	0.489	0.496	0.570	0.568	0.431	0.436	0.428	0.444	0.438	0.470	0.416	0.386	0.461
	0.82	0.552	0.541	0.516	0.519	0.432	0.439	0.497	0.496	0.397	0.400	0.394	0.406	0.400	0.429	0.376	0.346	0.424
	0.84	0.532	0.517	0.458	0.460	0.374	0.381	0.431	0.430	0.355	0.359	0.351	0.365	0.359	0.386	0.332	0.297	0.375
	0.86	0.445	0.432	0.386	0.388	0.308	0.313	0.365	0.364	0.308	0.312	0.307	0.318	0.313	0.337	0.276	0.245	0.313
	0.88	0.365	0.348	0.300	0.303	0.258	0.265	0.302	0.301	0.262	0.266	0.260	0.271	0.266	0.279	0.235	0.204	0.267
	0.90	0.296	0.284	0.241	0.243	0.205	0.210	0.227	0.227	0.219	0.220	0.217	0.225	0.222	0.229	0.192	0.172	0.219
	0.92	0.218	0.208	0.170	0.171	0.150	0.155	0.181	0.180	0.161	0.163	0.162	0.164	0.164	0.169	0.150	0.130	0.166
	0.94	0.138	0.129	0.122	0.123	0.123	0.127	0.127	0.126	0.117	0.119	0.115	0.119	0.119	0.121	0.112	0.098	0.121
	0.96	0.104	0.102	0.079	0.080	0.085	0.087	0.094	0.093	0.083	0.083	0.084	0.084	0.083	0.084	0.084	0.075	0.087
	0.98	0.081	0.076	0.062	0.062	0.059	0.060	0.066	0.065	0.066	0.066	0.066	0.067	0.067	0.068	0.066	0.061	0.067
200	0.80	0.873	0.873	0.923	0.926	0.711	0.901	0.930	0.930	0.685	0.683	0.684	0.687	0.682	0.760	0.691	0.684	0.751
	0.82	0.843	0.843	0.907	0.909	0.650	0.877	0.908	0.910	0.673	0.670	0.672	0.676	0.668	0.745	0.660	0.646	0.727
	0.84	0.834	0.836	0.871	0.875	0.578	0.844	0.877	0.878	0.663	0.661	0.663	0.665	0.659	0.723	0.639	0.618	0.714
	0.86	0.767	0.770	0.815	0.821	0.492	0.796	0.830	0.832	0.628	0.626	0.625	0.630	0.624	0.686	0.587	0.551	0.670
	0.88	0.701	0.703	0.742	0.748	0.388	0.719	0.756	0.760	0.588	0.585	0.586	0.590	0.583	0.636	0.527	0.481	0.622
	0.90	0.621	0.625	0.642	0.649	0.286	0.625	0.668	0.672	0.523	0.521	0.521	0.528	0.520	0.563	0.451	0.397	0.548
	0.92	0.465	0.468	0.498	0.507	0.183	0.494	0.524	0.530	0.431	0.429	0.428	0.434	0.427	0.465	0.353	0.298	0.447
	0.94	0.308	0.310	0.333	0.342	0.096	0.343	0.358	0.363	0.298	0.296	0.298	0.300	0.296	0.317	0.245	0.197	0.307
	0.96	0.186	0.187	0.187	0.191	0.050	0.184	0.203	0.206	0.170	0.168	0.168	0.171	0.168	0.179	0.137	0.114	0.169
	0.98	0.096	0.097	0.091	0.094	0.020	0.090	0.094	0.096	0.089	0.088	0.088	0.087	0.086	0.090	0.078	0.069	0.089
400	0.80	0.991	0.991	0.998	0.998	0.991	0.991	0.996	0.996	0.920	0.919	0.920	0.919	0.917	0.990	0.921	0.952	0.951
	0.82	0.986	0.986	0.997	0.997	0.990	0.990	0.995	0.995	0.923	0.923	0.921	0.923	0.920	0.987	0.915	0.940	0.949
	0.84	0.987	0.986	0.995	0.995	0.989	0.990	0.992	0.993	0.917	0.915	0.916	0.917	0.913	0.980	0.906	0.917	0.943
	0.86	0.980	0.979	0.990	0.990	0.985	0.985	0.986	0.986	0.918	0.916	0.917	0.917	0.914	0.973	0.903	0.897	0.941
	0.88	0.965	0.964	0.983	0.982	0.972	0.973	0.979	0.979	0.902	0.901	0.899	0.902	0.899	0.951	0.879	0.859	0.923
	0.90	0.949	0.946	0.962	0.961	0.955	0.956	0.961	0.962	0.888	0.887	0.886	0.889	0.885	0.928	0.853	0.813	0.909
	0.92	0.900	0.899	0.930	0.929	0.905	0.907	0.910	0.912	0.847	0.846	0.843	0.847	0.842	0.879	0.792	0.727	0.865
	0.94	0.812	0.811	0.814	0.812	0.803	0.807	0.801	0.804	0.747	0.748	0.740	0.751	0.745	0.775	0.659	0.566	0.758
	0.96	0.589	0.582	0.562	0.559	0.559	0.564	0.540	0.545	0.513	0.514	0.501	0.519	0.512	0.529	0.421	0.325	0.515
	0.98	0.228	0.218	0.208	0.207	0.205	0.208	0.198	0.201	0.198	0.197	0.191	0.200	0.196	0.198	0.158	0.119	0.191

Table 3.10. Power of the tests for model with an intercept only, $\phi = 0$.

N	ρ	BNM_0	$BEPO_0$	BNM_b	$BEPO_b$	BNM_a	$BEPO_a$	BNM_h	$BEPO_h$	MZ_a	MZ_t	MSB	P_t	MP_t	ADF	MZ_{aLS}	ADF_{LS}	MZ_{a2}
50	0.1	0.959	0.955	0.950	0.952	0.773	0.779	0.868	0.880	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.2	0.950	0.943	0.925	0.929	0.772	0.776	0.871	0.878	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.3	0.934	0.927	0.918	0.919	0.746	0.753	0.848	0.855	1.000	1.000	1.000	1.000	1.000	1.000	0.999	1.000	1.000
	0.4	0.917	0.907	0.894	0.898	0.725	0.731	0.824	0.833	1.000	1.000	1.000	1.000	1.000	1.000	0.994	0.999	1.000
	0.5	0.896	0.886	0.861	0.864	0.693	0.699	0.795	0.806	1.000	1.000	1.000	0.999	0.999	1.000	0.954	0.987	1.000
	0.6	0.854	0.840	0.817	0.826	0.610	0.619	0.727	0.741	0.997	0.996	0.994	0.993	0.994	0.999	0.810	0.921	0.997
	0.7	0.783	0.763	0.741	0.748	0.497	0.508	0.624	0.638	0.965	0.955	0.942	0.951	0.950	0.984	0.513	0.699	0.965
	0.8	0.550	0.566	0.530	0.542	0.336	0.338	0.408	0.424	0.772	0.753	0.713	0.737	0.748	0.862	0.216	0.383	0.769
	0.9	0.211	0.222	0.241	0.244	0.179	0.179	0.188	0.196	0.366	0.346	0.319	0.309	0.345	0.488	0.061	0.141	0.355
	1	0.054	0.044	0.053	0.053	0.047	0.047	0.047	0.047	0.062	0.056	0.058	0.044	0.056	0.096	0.008	0.057	0.057
100	0.1	0.999	0.998	0.991	0.992	0.941	0.947	0.975	0.978	0.715	0.715	0.702	0.716	0.714	0.893	0.698	0.742	0.779
	0.2	0.996	0.996	0.992	0.992	0.949	0.954	0.983	0.984	0.742	0.739	0.730	0.741	0.739	0.886	0.711	0.743	0.797
	0.3	0.994	0.994	0.990	0.991	0.950	0.951	0.981	0.983	0.756	0.754	0.744	0.756	0.753	0.883	0.725	0.744	0.811
	0.4	0.992	0.992	0.989	0.990	0.949	0.952	0.979	0.983	0.784	0.783	0.772	0.785	0.783	0.879	0.739	0.743	0.832
	0.5	0.987	0.988	0.983	0.985	0.955	0.958	0.977	0.979	0.802	0.800	0.790	0.801	0.799	0.874	0.755	0.744	0.850
	0.6	0.982	0.981	0.976	0.978	0.953	0.956	0.969	0.974	0.823	0.820	0.809	0.819	0.818	0.874	0.765	0.741	0.858
	0.7	0.966	0.965	0.961	0.965	0.928	0.933	0.957	0.963	0.834	0.830	0.820	0.830	0.830	0.868	0.749	0.710	0.868
	0.8	0.926	0.930	0.913	0.922	0.874	0.882	0.901	0.911	0.841	0.836	0.823	0.831	0.834	0.855	0.646	0.597	0.872
	0.9	0.626	0.631	0.607	0.625	0.553	0.567	0.609	0.632	0.632	0.614	0.572	0.588	0.610	0.660	0.236	0.225	0.651
	1	0.051	0.047	0.048	0.047	0.051	0.052	0.052	0.054	0.047	0.044	0.044	0.038	0.043	0.053	0.015	0.042	0.048
200	0.1	1.000	1.000	0.999	0.999	0.970	0.975	0.991	0.992	0.859	0.855	0.842	0.855	0.854	0.998	0.795	0.920	0.896
	0.2	1.000	1.000	0.998	0.999	0.978	0.980	0.993	0.993	0.882	0.880	0.868	0.881	0.880	0.997	0.819	0.910	0.911
	0.3	1.000	1.000	0.999	0.999	0.977	0.980	0.998	0.998	0.898	0.895	0.887	0.896	0.894	0.996	0.832	0.899	0.924
	0.4	1.000	1.000	1.000	1.000	0.978	0.981	0.996	0.996	0.918	0.915	0.906	0.914	0.913	0.995	0.844	0.893	0.939
	0.5	1.000	1.000	1.000	1.000	0.980	0.983	0.998	0.998	0.930	0.927	0.920	0.926	0.925	0.993	0.860	0.878	0.949
	0.6	1.000	1.000	0.999	0.999	0.986	0.987	0.997	0.997	0.946	0.941	0.935	0.941	0.941	0.991	0.873	0.867	0.959
	0.7	0.999	0.999	0.999	0.999	0.989	0.990	0.997	0.997	0.953	0.950	0.945	0.949	0.948	0.987	0.885	0.862	0.964
	0.8	0.999	0.999	0.997	0.997	0.986	0.987	0.996	0.996	0.952	0.949	0.944	0.947	0.948	0.970	0.887	0.842	0.964
	0.9	0.967	0.969	0.963	0.969	0.940	0.947	0.963	0.970	0.926	0.921	0.911	0.915	0.919	0.927	0.754	0.644	0.939
	1	0.050	0.047	0.048	0.046	0.048	0.050	0.056	0.055	0.051	0.046	0.046	0.042	0.046	0.050	0.022	0.046	0.052
400	0.1	1.000	1.000	1.000	1.000	0.985	0.986	0.998	0.998	0.982	0.981	0.979	0.979	0.979	1.000	0.942	1.000	0.987
	0.2	1.000	1.000	1.000	1.000	0.988	0.990	0.999	0.999	0.990	0.989	0.988	0.989	0.989	1.000	0.958	1.000	0.994
	0.3	1.000	1.000	1.000	1.000	0.990	0.991	0.998	0.999	0.994	0.992	0.992	0.991	0.991	1.000	0.969	1.000	0.995
	0.4	1.000	1.000	1.000	1.000	0.993	0.993	0.998	0.998	0.996	0.996	0.996	0.995	0.995	1.000	0.980	1.000	0.998
	0.5	1.000	1.000	1.000	1.000	0.992	0.993	0.999	0.999	0.998	0.998	0.998	0.997	0.997	1.000	0.983	0.999	0.999
	0.6	1.000	1.000	1.000	1.000	0.994	0.994	1.000	1.000	1.000	0.999	1.000	0.999	0.999	1.000	0.990	1.000	1.000
	0.7	1.000	1.000	1.000	1.000	0.998	0.997	1.000	1.000	1.000	1.000	0.999	0.999	0.999	1.000	0.992	0.997	1.000
	0.8	1.000	1.000	1.000	1.000	0.999	0.999	1.000	1.000	1.000	0.999	0.999	0.999	0.999	1.000	0.991	0.987	1.000
	0.9	0.999	0.999	1.000	1.000	0.999	0.999	1.000	1.000	0.997	0.996	0.997	0.995	0.995	0.998	0.974	0.945	0.998
	1	0.047	0.045	0.054	0.055	0.043	0.041	0.048	0.046	0.047	0.043	0.043	0.040	0.043	0.044	0.025	0.042	0.047

Table 3.11. Power of the tests for model with an intercept and a trend, $\phi = 0$.

N	ρ	BNM_0	$BEPO_0$	BNM_b	$BEPO_b$	BNM_a	$BEPO_a$	BNM_h	$BEPO_h$	MZ_a	MZ_t	MSB	P_t	MP_t	ADF	MZ_{aLS}	ADF_{LS}	MZ_{a2}
50	0.1	0.972	0.957	0.924	0.923	0.661	0.669	0.829	0.830	1.000	1.000	1.000	0.999	0.999	1.000	0.985	1.000	1.000
	0.2	0.953	0.929	0.906	0.905	0.651	0.661	0.802	0.802	0.999	0.998	0.999	0.997	0.996	1.000	0.941	1.000	0.999
	0.3	0.929	0.895	0.873	0.871	0.626	0.639	0.794	0.795	0.990	0.986	0.989	0.988	0.983	1.000	0.808	0.999	0.989
	0.4	0.895	0.846	0.823	0.820	0.557	0.572	0.734	0.734	0.948	0.944	0.944	0.956	0.940	0.999	0.574	0.990	0.944
	0.5	0.827	0.759	0.753	0.750	0.501	0.517	0.663	0.663	0.804	0.803	0.800	0.833	0.801	0.992	0.308	0.938	0.794
	0.6	0.719	0.626	0.624	0.620	0.390	0.406	0.548	0.548	0.544	0.544	0.540	0.598	0.556	0.946	0.130	0.786	0.526
	0.7	0.533	0.424	0.429	0.424	0.266	0.278	0.371	0.370	0.261	0.265	0.258	0.299	0.277	0.771	0.038	0.519	0.243
	0.8	0.272	0.275	0.236	0.234	0.173	0.173	0.213	0.213	0.092	0.095	0.095	0.106	0.101	0.468	0.009	0.262	0.082
	0.9	0.092	0.092	0.093	0.091	0.090	0.090	0.094	0.093	0.029	0.029	0.030	0.032	0.031	0.209	0.002	0.125	0.025
	1	0.055	0.035	0.048	0.047	0.053	0.054	0.051	0.051	0.010	0.010	0.010	0.011	0.011	0.090	0.001	0.068	0.009
100	0.1	0.999	0.999	0.994	0.994	0.866	0.865	0.959	0.959	0.605	0.606	0.604	0.610	0.607	0.687	0.723	0.760	0.759
	0.2	0.999	0.998	0.994	0.994	0.876	0.875	0.963	0.962	0.642	0.641	0.642	0.646	0.642	0.710	0.729	0.756	0.769
	0.3	0.997	0.996	0.991	0.991	0.867	0.866	0.964	0.963	0.670	0.670	0.668	0.675	0.671	0.722	0.742	0.759	0.785
	0.4	0.994	0.992	0.987	0.987	0.877	0.877	0.954	0.954	0.697	0.696	0.696	0.702	0.697	0.738	0.745	0.751	0.791
	0.5	0.983	0.978	0.978	0.978	0.867	0.866	0.948	0.947	0.720	0.720	0.719	0.725	0.721	0.748	0.750	0.743	0.800
	0.6	0.968	0.959	0.956	0.956	0.847	0.846	0.918	0.918	0.736	0.737	0.734	0.741	0.738	0.759	0.739	0.732	0.803
	0.7	0.911	0.889	0.896	0.895	0.788	0.786	0.847	0.846	0.700	0.700	0.698	0.705	0.701	0.735	0.635	0.671	0.760
	0.8	0.684	0.683	0.703	0.705	0.572	0.569	0.628	0.627	0.505	0.509	0.498	0.519	0.515	0.632	0.339	0.457	0.546
	0.9	0.281	0.273	0.278	0.281	0.220	0.218	0.230	0.228	0.133	0.135	0.132	0.137	0.139	0.250	0.069	0.146	0.145
	1	0.052	0.042	0.051	0.051	0.055	0.054	0.050	0.049	0.020	0.021	0.020	0.021	0.022	0.043	0.010	0.042	0.021
200	0.1	1.000	1.000	0.998	0.998	0.945	0.946	0.988	0.988	0.634	0.636	0.633	0.640	0.638	0.827	0.791	0.847	0.820
	0.2	1.000	1.000	0.998	0.998	0.949	0.949	0.992	0.992	0.684	0.685	0.683	0.688	0.686	0.840	0.811	0.854	0.844
	0.3	1.000	1.000	0.999	0.999	0.952	0.952	0.992	0.992	0.726	0.725	0.725	0.728	0.727	0.844	0.825	0.847	0.856
	0.4	1.000	1.000	1.000	1.000	0.955	0.955	0.991	0.991	0.762	0.762	0.762	0.763	0.762	0.849	0.830	0.840	0.863
	0.5	1.000	1.000	1.000	1.000	0.956	0.957	0.993	0.993	0.791	0.791	0.790	0.793	0.791	0.847	0.843	0.834	0.876
	0.6	0.999	0.999	0.999	0.999	0.967	0.967	0.996	0.996	0.814	0.813	0.814	0.814	0.813	0.848	0.852	0.834	0.888
	0.7	0.995	0.995	0.996	0.996	0.974	0.974	0.994	0.994	0.838	0.839	0.837	0.840	0.839	0.851	0.859	0.829	0.893
	0.8	0.964	0.964	0.968	0.968	0.947	0.948	0.967	0.965	0.844	0.844	0.844	0.845	0.845	0.845	0.844	0.800	0.890
	0.9	0.718	0.718	0.722	0.722	0.660	0.662	0.710	0.705	0.603	0.603	0.601	0.603	0.608	0.654	0.470	0.457	0.634
	1	0.050	0.044	0.056	0.056	0.047	0.048	0.049	0.048	0.027	0.027	0.028	0.027	0.028	0.039	0.020	0.042	0.028
400	0.1	1.000	1.000	0.999	0.999	0.972	0.972	0.997	0.997	0.785	0.785	0.785	0.786	0.785	0.991	0.908	0.997	0.937
	0.2	1.000	1.000	1.000	1.000	0.976	0.975	0.998	0.998	0.836	0.835	0.837	0.835	0.835	0.995	0.929	0.998	0.953
	0.3	1.000	1.000	1.000	1.000	0.977	0.977	0.997	0.997	0.879	0.877	0.881	0.877	0.876	0.996	0.937	0.994	0.963
	0.4	1.000	1.000	1.000	1.000	0.983	0.982	0.998	0.998	0.918	0.916	0.921	0.916	0.916	0.998	0.954	0.994	0.973
	0.5	1.000	1.000	1.000	1.000	0.984	0.983	0.999	0.999	0.941	0.939	0.942	0.939	0.938	0.998	0.957	0.990	0.976
	0.6	1.000	1.000	1.000	1.000	0.990	0.990	0.999	0.999	0.966	0.966	0.966	0.966	0.965	0.998	0.969	0.986	0.985
	0.7	1.000	1.000	1.000	1.000	0.996	0.996	1.000	1.000	0.974	0.974	0.975	0.974	0.973	0.995	0.971	0.971	0.986
	0.8	1.000	1.000	0.999	0.999	0.997	0.997	1.000	1.000	0.974	0.973	0.975	0.973	0.973	0.984	0.965	0.943	0.986
	0.9	0.980	0.980	0.977	0.977	0.971	0.971	0.975	0.975	0.948	0.947	0.948	0.947	0.947	0.942	0.931	0.883	0.963
	1	0.047	0.044	0.048	0.046	0.050	0.049	0.047	0.047	0.036	0.035	0.037	0.034	0.036	0.040	0.028	0.043	0.036

Table 3.12. Empirical size of the tests for model with an intercept only.

	ϕ	$\overline{BNM_b}$	$\overline{BEPO_b}$	$\overline{BNM_a}$	$\overline{BEPO_a}$	$\overline{BNM_h}$	$\overline{BEPO_h}$	ADF_{PQ}	ADF_S	M_{PQ}	M_S
50	−0.8	0.522	0.511	0.508	0.496	0.500	0.488	0.246	0.565	0.161	0.481
	−0.7	0.381	0.372	0.367	0.355	0.359	0.348	0.175	0.456	0.103	0.388
	−0.6	0.291	0.280	0.277	0.266	0.273	0.263	0.141	0.375	0.074	0.321
	−0.5	0.235	0.222	0.232	0.221	0.224	0.212	0.120	0.325	0.070	0.273
	−0.4	0.207	0.194	0.205	0.196	0.194	0.183	0.116	0.303	0.071	0.240
	−0.3	0.175	0.156	0.182	0.167	0.173	0.156	0.112	0.259	0.071	0.192
	−0.2	0.130	0.115	0.155	0.142	0.137	0.123	0.105	0.208	0.068	0.145
	−0.1	0.092	0.078	0.132	0.121	0.110	0.096	0.088	0.141	0.056	0.093
	0.0	0.069	0.058	0.115	0.104	0.088	0.076	0.075	0.104	0.048	0.071
	0.1	0.061	0.051	0.113	0.102	0.081	0.071	0.053	0.079	0.035	0.059
	0.2	0.059	0.051	0.117	0.108	0.089	0.078	0.042	0.064	0.029	0.057
	0.3	0.059	0.050	0.109	0.098	0.082	0.072	0.045	0.059	0.037	0.069
	0.4	0.069	0.061	0.114	0.103	0.090	0.081	0.047	0.053	0.046	0.074
	0.5	0.069	0.063	0.109	0.099	0.089	0.080	0.048	0.056	0.054	0.088
	0.6	0.073	0.060	0.114	0.102	0.091	0.078	0.051	0.059	0.062	0.097
	0.7	0.066	0.058	0.109	0.098	0.085	0.076	0.050	0.066	0.070	0.108
	0.8	0.063	0.052	0.097	0.085	0.077	0.066	0.045	0.075	0.072	0.117
100	−0.8	0.259	0.256	0.256	0.252	0.246	0.241	0.112	0.276	0.042	0.140
	−0.7	0.147	0.140	0.162	0.155	0.146	0.138	0.099	0.152	0.042	0.083
	−0.6	0.113	0.106	0.129	0.121	0.116	0.109	0.084	0.097	0.043	0.064
	−0.5	0.101	0.095	0.111	0.105	0.103	0.098	0.081	0.080	0.051	0.063
	−0.4	0.105	0.100	0.101	0.095	0.094	0.088	0.074	0.096	0.051	0.086
	−0.3	0.118	0.112	0.107	0.102	0.106	0.099	0.068	0.131	0.053	0.115
	−0.2	0.116	0.111	0.103	0.096	0.105	0.099	0.069	0.133	0.057	0.116
	−0.1	0.079	0.073	0.081	0.075	0.077	0.072	0.070	0.115	0.058	0.095
	0.0	0.057	0.049	0.076	0.068	0.064	0.055	0.055	0.071	0.044	0.057
	0.1	0.046	0.042	0.069	0.065	0.056	0.052	0.041	0.052	0.033	0.043
	0.2	0.047	0.042	0.073	0.066	0.058	0.051	0.041	0.048	0.037	0.050
	0.3	0.057	0.054	0.075	0.071	0.066	0.061	0.044	0.047	0.042	0.057
	0.4	0.056	0.051	0.072	0.066	0.064	0.059	0.057	0.052	0.059	0.064
	0.5	0.058	0.054	0.072	0.066	0.062	0.058	0.045	0.055	0.052	0.067
	0.6	0.054	0.049	0.069	0.061	0.058	0.052	0.045	0.057	0.052	0.072
	0.7	0.061	0.056	0.075	0.070	0.068	0.061	0.042	0.059	0.056	0.078
	0.8	0.062	0.056	0.076	0.070	0.067	0.060	0.041	0.063	0.062	0.084
200	−0.8	0.105	0.102	0.114	0.113	0.107	0.104	0.081	0.203	0.018	0.052
	−0.7	0.085	0.081	0.093	0.090	0.086	0.083	0.071	0.103	0.028	0.040
	−0.6	0.070	0.068	0.080	0.078	0.073	0.071	0.072	0.067	0.043	0.039
	−0.5	0.070	0.067	0.076	0.072	0.071	0.068	0.059	0.050	0.043	0.038
	−0.4	0.066	0.063	0.073	0.070	0.069	0.065	0.058	0.051	0.050	0.047
	−0.3	0.061	0.059	0.064	0.062	0.059	0.057	0.057	0.055	0.049	0.053
	−0.2	0.082	0.082	0.070	0.068	0.072	0.071	0.054	0.083	0.050	0.080
	−0.1	0.076	0.073	0.068	0.065	0.071	0.068	0.053	0.085	0.050	0.078
	0.0	0.053	0.050	0.059	0.057	0.054	0.051	0.046	0.056	0.042	0.050
	0.1	0.041	0.039	0.056	0.053	0.048	0.046	0.043	0.045	0.039	0.042
	0.2	0.053	0.050	0.064	0.061	0.059	0.056	0.048	0.046	0.046	0.048
	0.3	0.055	0.052	0.063	0.060	0.057	0.055	0.050	0.050	0.050	0.056
	0.4	0.057	0.054	0.062	0.060	0.058	0.056	0.045	0.052	0.049	0.057
	0.5	0.057	0.057	0.062	0.062	0.058	0.058	0.042	0.050	0.047	0.056
	0.6	0.051	0.049	0.057	0.054	0.053	0.051	0.043	0.052	0.049	0.059
	0.7	0.049	0.045	0.056	0.051	0.052	0.047	0.041	0.054	0.052	0.062
	0.8	0.058	0.054	0.064	0.059	0.060	0.056	0.038	0.052	0.053	0.061
400	−0.8	0.078	0.076	0.081	0.080	0.078	0.077	0.071	0.165	0.017	0.038
	−0.7	0.068	0.067	0.071	0.070	0.069	0.068	0.067	0.087	0.036	0.038
	−0.6	0.066	0.063	0.070	0.066	0.067	0.064	0.061	0.060	0.044	0.040
	−0.5	0.060	0.058	0.063	0.061	0.061	0.060	0.054	0.052	0.045	0.044
	−0.4	0.058	0.056	0.061	0.059	0.059	0.057	0.053	0.048	0.049	0.045
	−0.3	0.054	0.052	0.059	0.057	0.056	0.053	0.054	0.049	0.052	0.048
	−0.2	0.059	0.059	0.058	0.059	0.056	0.057	0.048	0.054	0.046	0.054
	−0.1	0.072	0.069	0.061	0.060	0.064	0.062	0.053	0.074	0.051	0.073
	0.0	0.048	0.046	0.052	0.049	0.049	0.047	0.050	0.055	0.047	0.053
	0.1	0.044	0.043	0.053	0.052	0.049	0.047	0.047	0.043	0.047	0.043
	0.2	0.050	0.047	0.055	0.051	0.052	0.049	0.051	0.051	0.051	0.052
	0.3	0.051	0.049	0.054	0.052	0.052	0.050	0.046	0.048	0.047	0.050
	0.4	0.051	0.050	0.055	0.053	0.052	0.051	0.042	0.046	0.044	0.048
	0.5	0.052	0.050	0.055	0.052	0.053	0.051	0.047	0.053	0.050	0.056
	0.6	0.056	0.054	0.060	0.057	0.056	0.055	0.042	0.048	0.047	0.051
	0.7	0.054	0.053	0.058	0.055	0.055	0.053	0.045	0.051	0.051	0.055
	0.8	0.051	0.048	0.056	0.053	0.052	0.050	0.044	0.050	0.054	0.055

Table 3.13. Empirical size of the tests for model with an intercept and a trend.

N	ϕ	$\overline{BNM}_b$	$\overline{BEPO}_b$	$\overline{BNM}_a$	$\overline{BEPO}_a$	$\overline{BNM}_h$	$\overline{BEPO}_h$	ADF_{PQ}	ADF_S	M_{PQ}	M_S
50	−0.8	0.775	0.729	0.679	0.635	0.738	0.690	0.390	0.874	0.266	0.801
	−0.7	0.671	0.623	0.582	0.540	0.632	0.586	0.268	0.792	0.156	0.678
	−0.6	0.548	0.500	0.472	0.431	0.503	0.458	0.202	0.712	0.100	0.526
	−0.5	0.445	0.387	0.389	0.340	0.410	0.358	0.161	0.613	0.066	0.359
	−0.4	0.329	0.270	0.301	0.251	0.310	0.256	0.150	0.524	0.053	0.229
	−0.3	0.241	0.193	0.245	0.203	0.241	0.195	0.136	0.403	0.036	0.131
	−0.2	0.162	0.123	0.181	0.144	0.166	0.127	0.119	0.278	0.025	0.067
	−0.1	0.102	0.072	0.149	0.115	0.120	0.090	0.094	0.177	0.013	0.038
	0.0	0.071	0.051	0.117	0.094	0.095	0.073	0.070	0.115	0.009	0.036
	0.1	0.064	0.042	0.125	0.096	0.094	0.068	0.040	0.072	0.004	0.043
	0.2	0.066	0.048	0.120	0.093	0.096	0.072	0.029	0.057	0.004	0.055
	0.3	0.076	0.053	0.125	0.095	0.097	0.071	0.017	0.037	0.005	0.066
	0.4	0.083	0.058	0.126	0.097	0.107	0.079	0.016	0.027	0.012	0.075
	0.5	0.089	0.065	0.130	0.101	0.108	0.080	0.016	0.027	0.019	0.087
	0.6	0.073	0.048	0.113	0.085	0.090	0.064	0.018	0.036	0.029	0.100
	0.7	0.082	0.058	0.117	0.092	0.098	0.074	0.019	0.051	0.034	0.133
	0.8	0.076	0.051	0.115	0.083	0.093	0.065	0.016	0.065	0.042	0.156
100	−0.8	0.601	0.588	0.515	0.499	0.540	0.527	0.171	0.516	0.093	0.449
	−0.7	0.361	0.347	0.332	0.312	0.323	0.304	0.116	0.320	0.050	0.283
	−0.6	0.237	0.217	0.245	0.223	0.226	0.204	0.087	0.211	0.039	0.197
	−0.5	0.199	0.184	0.192	0.173	0.176	0.159	0.083	0.196	0.039	0.179
	−0.4	0.184	0.165	0.167	0.150	0.156	0.137	0.072	0.224	0.036	0.181
	−0.3	0.188	0.167	0.161	0.143	0.161	0.142	0.063	0.232	0.035	0.163
	−0.2	0.136	0.120	0.137	0.120	0.130	0.113	0.060	0.195	0.031	0.115
	−0.1	0.095	0.077	0.114	0.095	0.100	0.082	0.059	0.124	0.029	0.064
	0.0	0.062	0.050	0.099	0.084	0.076	0.062	0.042	0.067	0.019	0.036
	0.1	0.049	0.040	0.094	0.079	0.066	0.055	0.028	0.042	0.012	0.029
	0.2	0.047	0.041	0.087	0.074	0.063	0.053	0.021	0.032	0.010	0.043
	0.3	0.068	0.055	0.110	0.095	0.081	0.069	0.023	0.025	0.018	0.050
	0.4	0.066	0.055	0.097	0.084	0.077	0.065	0.033	0.028	0.032	0.058
	0.5	0.061	0.050	0.092	0.078	0.072	0.060	0.031	0.029	0.036	0.066
	0.6	0.064	0.050	0.094	0.082	0.076	0.062	0.026	0.036	0.036	0.069
	0.7	0.064	0.050	0.101	0.083	0.077	0.061	0.021	0.045	0.035	0.083
	0.8	0.060	0.046	0.092	0.076	0.071	0.057	0.018	0.047	0.047	0.094
200	−0.8	0.199	0.190	0.219	0.207	0.200	0.189	0.086	0.203	0.021	0.084
	−0.7	0.130	0.121	0.153	0.145	0.137	0.128	0.062	0.068	0.019	0.031
	−0.6	0.103	0.096	0.121	0.114	0.108	0.102	0.056	0.034	0.025	0.023
	−0.5	0.083	0.075	0.099	0.091	0.089	0.080	0.055	0.026	0.032	0.029
	−0.4	0.079	0.072	0.095	0.088	0.081	0.074	0.045	0.027	0.032	0.032
	−0.3	0.082	0.076	0.085	0.078	0.074	0.068	0.049	0.067	0.038	0.065
	−0.2	0.110	0.102	0.091	0.082	0.094	0.087	0.044	0.109	0.034	0.091
	−0.1	0.091	0.082	0.086	0.076	0.084	0.075	0.050	0.104	0.037	0.077
	0.0	0.048	0.044	0.063	0.058	0.052	0.048	0.038	0.056	0.024	0.039
	0.1	0.043	0.039	0.070	0.063	0.054	0.049	0.024	0.034	0.018	0.031
	0.2	0.051	0.045	0.070	0.064	0.059	0.054	0.030	0.034	0.025	0.042
	0.3	0.057	0.051	0.070	0.064	0.061	0.055	0.038	0.032	0.038	0.046
	0.4	0.056	0.049	0.067	0.060	0.058	0.050	0.035	0.033	0.038	0.047
	0.5	0.048	0.044	0.064	0.059	0.054	0.050	0.027	0.036	0.033	0.051
	0.6	0.054	0.049	0.067	0.060	0.057	0.051	0.024	0.040	0.035	0.055
	0.7	0.059	0.053	0.072	0.065	0.064	0.058	0.024	0.043	0.044	0.064
	0.8	0.054	0.050	0.068	0.062	0.059	0.054	0.025	0.047	0.051	0.068
400	−0.8	0.108	0.104	0.118	0.114	0.110	0.106	0.061	0.187	0.007	0.025
	−0.7	0.083	0.079	0.091	0.088	0.085	0.082	0.055	0.071	0.017	0.021
	−0.6	0.073	0.069	0.084	0.080	0.076	0.072	0.049	0.046	0.027	0.025
	−0.5	0.066	0.063	0.073	0.070	0.068	0.065	0.045	0.034	0.035	0.027
	−0.4	0.064	0.061	0.073	0.070	0.067	0.063	0.046	0.033	0.038	0.033
	−0.3	0.060	0.057	0.068	0.064	0.062	0.059	0.042	0.032	0.037	0.035
	−0.2	0.066	0.062	0.063	0.060	0.061	0.058	0.044	0.054	0.039	0.055
	−0.1	0.081	0.078	0.068	0.065	0.071	0.067	0.036	0.074	0.033	0.066
	0.0	0.051	0.049	0.056	0.054	0.052	0.050	0.039	0.050	0.034	0.043
	0.1	0.041	0.039	0.057	0.055	0.050	0.047	0.034	0.038	0.030	0.037
	0.2	0.052	0.050	0.063	0.060	0.055	0.053	0.039	0.039	0.040	0.045
	0.3	0.055	0.052	0.062	0.059	0.057	0.054	0.036	0.038	0.038	0.046
	0.4	0.053	0.049	0.059	0.056	0.055	0.052	0.030	0.037	0.032	0.043
	0.5	0.049	0.047	0.054	0.051	0.051	0.048	0.030	0.038	0.036	0.047
	0.6	0.052	0.049	0.058	0.055	0.052	0.049	0.032	0.038	0.041	0.044
	0.7	0.053	0.050	0.058	0.055	0.053	0.051	0.028	0.039	0.041	0.047
	0.8	0.052	0.050	0.057	0.054	0.053	0.050	0.028	0.044	0.051	0.053

Table 3.14. Power of the tests for model with an intercept only, $\phi = 0$.

N	ρ	$\overline{BNM}_b$	$\overline{BEPO}_b$	$\overline{BNM}_a$	$\overline{BEPO}_a$	$\overline{BNM}_h$	$\overline{BEPO}_h$	ADF_{PQ}	ADF_S	M_{PQ}	M_S
50	0.80	0.603	0.576	0.655	0.633	0.625	0.601	0.630	0.831	0.501	0.698
	0.82	0.557	0.522	0.617	0.589	0.581	0.548	0.587	0.772	0.454	0.629
	0.84	0.497	0.466	0.564	0.540	0.525	0.497	0.525	0.707	0.394	0.561
	0.86	0.427	0.395	0.505	0.477	0.461	0.431	0.465	0.630	0.333	0.483
	0.88	0.366	0.334	0.455	0.427	0.404	0.374	0.397	0.543	0.273	0.402
	0.90	0.305	0.278	0.396	0.372	0.342	0.316	0.345	0.474	0.225	0.335
	0.92	0.242	0.218	0.337	0.316	0.278	0.256	0.281	0.385	0.182	0.269
	0.94	0.185	0.165	0.269	0.251	0.219	0.201	0.210	0.289	0.128	0.194
	0.96	0.139	0.121	0.220	0.202	0.169	0.151	0.159	0.222	0.095	0.146
	0.98	0.101	0.087	0.164	0.150	0.126	0.112	0.110	0.153	0.066	0.101
100	0.80	0.924	0.922	0.912	0.911	0.921	0.919	0.818	0.983	0.795	0.974
	0.82	0.902	0.900	0.891	0.889	0.897	0.896	0.807	0.974	0.777	0.963
	0.84	0.874	0.872	0.869	0.866	0.870	0.868	0.790	0.959	0.756	0.940
	0.86	0.824	0.823	0.823	0.821	0.822	0.821	0.759	0.930	0.721	0.901
	0.88	0.760	0.755	0.767	0.762	0.763	0.757	0.708	0.874	0.658	0.829
	0.90	0.644	0.640	0.664	0.659	0.649	0.644	0.622	0.782	0.565	0.720
	0.92	0.509	0.500	0.543	0.534	0.519	0.509	0.505	0.638	0.443	0.568
	0.94	0.369	0.360	0.411	0.403	0.382	0.373	0.366	0.467	0.309	0.401
	0.96	0.233	0.222	0.273	0.263	0.247	0.237	0.236	0.308	0.197	0.260
	0.98	0.121	0.115	0.156	0.149	0.133	0.126	0.129	0.169	0.103	0.137
200	0.80	0.997	0.998	0.991	0.992	0.996	0.997	0.917	0.999	0.893	0.999
	0.82	0.997	0.998	0.991	0.992	0.995	0.996	0.923	0.998	0.903	0.998
	0.84	0.995	0.995	0.988	0.988	0.993	0.994	0.924	0.998	0.909	0.998
	0.86	0.991	0.992	0.982	0.983	0.990	0.991	0.923	0.999	0.912	0.998
	0.88	0.984	0.985	0.977	0.978	0.983	0.984	0.914	0.996	0.904	0.994
	0.90	0.972	0.974	0.964	0.966	0.970	0.971	0.897	0.989	0.890	0.986
	0.92	0.925	0.930	0.914	0.919	0.922	0.926	0.857	0.960	0.847	0.952
	0.94	0.796	0.802	0.792	0.796	0.795	0.801	0.745	0.855	0.723	0.832
	0.96	0.541	0.543	0.555	0.558	0.546	0.548	0.509	0.602	0.482	0.569
	0.98	0.233	0.236	0.255	0.255	0.238	0.240	0.211	0.256	0.193	0.234
400	0.80	1.000	1.000	0.999	0.999	1.000	1.000	0.975	1.000	0.960	1.000
	0.82	1.000	1.000	0.999	0.999	1.000	1.000	0.976	1.000	0.964	1.000
	0.84	1.000	1.000	1.000	1.000	1.000	1.000	0.982	1.000	0.973	1.000
	0.86	1.000	1.000	0.999	0.999	1.000	1.000	0.984	1.000	0.977	1.000
	0.88	1.000	1.000	0.999	1.000	1.000	1.000	0.985	1.000	0.979	1.000
	0.90	1.000	1.000	0.999	0.999	1.000	1.000	0.987	1.000	0.982	1.000
	0.92	0.999	1.000	0.999	0.999	0.999	1.000	0.985	1.000	0.980	1.000
	0.94	0.998	0.998	0.995	0.995	0.997	0.998	0.975	0.998	0.973	0.998
	0.96	0.954	0.959	0.948	0.953	0.953	0.957	0.917	0.968	0.913	0.964
	0.98	0.554	0.569	0.563	0.578	0.556	0.570	0.515	0.572	0.506	0.559

Table 3.15. Power of the tests for model with an intercept and a trend, $\phi = 0$.

N	ρ	$\overline{BNM}_b$	$\overline{BEPO}_b$	$\overline{BNM}_a$	$\overline{BEPO}_a$	$\overline{BNM}_h$	$\overline{BEPO}_h$	ADF_{PQ}	ADF_S	M_{PQ}	M_S
50	0.80	0.323	0.239	0.372	0.294	0.350	0.268	0.322	0.465	0.058	0.137
	0.82	0.286	0.206	0.340	0.268	0.313	0.237	0.283	0.421	0.044	0.115
	0.84	0.241	0.176	0.300	0.237	0.270	0.204	0.239	0.360	0.034	0.099
	0.86	0.220	0.158	0.282	0.222	0.248	0.186	0.200	0.305	0.028	0.080
	0.88	0.179	0.130	0.245	0.197	0.208	0.160	0.170	0.266	0.022	0.074
	0.90	0.149	0.105	0.208	0.162	0.176	0.131	0.140	0.222	0.019	0.063
	0.92	0.131	0.092	0.193	0.150	0.157	0.117	0.117	0.186	0.012	0.051
	0.94	0.103	0.073	0.162	0.125	0.128	0.095	0.096	0.158	0.012	0.047
	0.96	0.088	0.059	0.139	0.106	0.111	0.080	0.080	0.131	0.010	0.040
	0.98	0.081	0.057	0.133	0.104	0.103	0.076	0.076	0.121	0.008	0.037
100	0.80	0.733	0.685	0.740	0.699	0.736	0.692	0.633	0.883	0.486	0.720
	0.82	0.666	0.615	0.690	0.646	0.675	0.627	0.582	0.810	0.420	0.618
	0.84	0.600	0.546	0.634	0.588	0.612	0.561	0.510	0.724	0.338	0.515
	0.86	0.490	0.436	0.543	0.493	0.507	0.454	0.424	0.613	0.258	0.402
	0.88	0.415	0.364	0.472	0.425	0.432	0.383	0.334	0.487	0.192	0.305
	0.90	0.308	0.264	0.373	0.331	0.332	0.288	0.249	0.375	0.132	0.216
	0.92	0.218	0.183	0.289	0.252	0.241	0.205	0.180	0.268	0.087	0.143
	0.94	0.157	0.134	0.220	0.192	0.178	0.154	0.124	0.192	0.056	0.099
	0.96	0.102	0.083	0.160	0.137	0.120	0.100	0.080	0.123	0.035	0.064
	0.98	0.069	0.056	0.107	0.091	0.080	0.067	0.059	0.089	0.025	0.043
200	0.80	0.975	0.971	0.967	0.963	0.973	0.969	0.862	0.996	0.854	0.997
	0.82	0.960	0.953	0.954	0.948	0.958	0.952	0.849	0.996	0.842	0.995
	0.84	0.940	0.931	0.933	0.925	0.938	0.929	0.829	0.992	0.816	0.986
	0.86	0.901	0.891	0.896	0.887	0.901	0.891	0.807	0.981	0.788	0.966
	0.88	0.843	0.825	0.842	0.826	0.843	0.826	0.754	0.945	0.718	0.904
	0.90	0.745	0.723	0.757	0.738	0.749	0.727	0.656	0.849	0.595	0.774
	0.92	0.589	0.559	0.615	0.589	0.597	0.568	0.507	0.665	0.431	0.566
	0.94	0.401	0.372	0.442	0.416	0.411	0.383	0.316	0.438	0.250	0.351
	0.96	0.215	0.196	0.249	0.232	0.221	0.204	0.154	0.223	0.115	0.166
	0.98	0.095	0.085	0.119	0.109	0.102	0.092	0.067	0.098	0.047	0.070
400	0.80	1.000	1.000	0.998	0.997	0.999	0.999	0.967	1.000	0.950	1.000
	0.82	0.999	0.999	0.998	0.998	0.999	0.998	0.964	1.000	0.950	1.000
	0.84	0.998	0.998	0.996	0.996	0.998	0.998	0.960	1.000	0.949	1.000
	0.86	0.996	0.996	0.994	0.994	0.995	0.995	0.954	1.000	0.951	1.000
	0.88	0.992	0.991	0.989	0.988	0.991	0.991	0.951	1.000	0.949	1.000
	0.90	0.978	0.976	0.975	0.973	0.978	0.976	0.939	1.000	0.940	1.000
	0.92	0.945	0.942	0.943	0.940	0.945	0.942	0.907	0.996	0.909	0.995
	0.94	0.851	0.843	0.851	0.844	0.852	0.845	0.819	0.940	0.806	0.921
	0.96	0.590	0.576	0.599	0.585	0.591	0.577	0.522	0.639	0.495	0.595
	0.98	0.215	0.207	0.233	0.223	0.218	0.209	0.158	0.205	0.139	0.178

Table 3.16. Power of the tests for model with an intercept only, $\phi = 0$.

N	ρ	$\overline{BNM}_b$	$\overline{BEPO}_b$	$\overline{BNM}_a$	$\overline{BEPO}_a$	$\overline{BNM}_h$	$\overline{BEPO}_h$	ADF_{PQ}	ADF_S	M_{PQ}	M_S
50	0.1	0.952	0.947	0.901	0.895	0.934	0.928	0.716	0.928	0.624	0.900
	0.2	0.943	0.939	0.898	0.893	0.925	0.921	0.725	0.945	0.645	0.923
	0.3	0.932	0.924	0.884	0.877	0.914	0.907	0.739	0.957	0.667	0.943
	0.4	0.909	0.901	0.874	0.866	0.895	0.887	0.749	0.966	0.681	0.956
	0.5	0.904	0.896	0.880	0.873	0.894	0.885	0.756	0.975	0.695	0.967
	0.6	0.850	0.836	0.844	0.832	0.850	0.836	0.753	0.976	0.694	0.960
	0.7	0.773	0.754	0.782	0.765	0.774	0.756	0.739	0.954	0.663	0.902
	0.8	0.603	0.576	0.655	0.633	0.625	0.601	0.622	0.827	0.495	0.699
	0.9	0.305	0.278	0.396	0.372	0.342	0.316	0.342	0.465	0.228	0.331
	1.0	0.069	0.058	0.115	0.104	0.088	0.076	0.075	0.102	0.047	0.067
100	0.1	0.993	0.993	0.960	0.962	0.982	0.983	0.728	0.812	0.630	0.727
	0.2	0.993	0.992	0.963	0.963	0.984	0.984	0.740	0.843	0.648	0.797
	0.3	0.991	0.991	0.963	0.962	0.985	0.985	0.757	0.898	0.677	0.874
	0.4	0.990	0.990	0.964	0.964	0.982	0.983	0.771	0.933	0.703	0.923
	0.5	0.988	0.988	0.963	0.964	0.980	0.980	0.792	0.962	0.733	0.957
	0.6	0.982	0.981	0.962	0.961	0.975	0.975	0.810	0.979	0.764	0.977
	0.7	0.968	0.966	0.956	0.955	0.964	0.963	0.826	0.986	0.793	0.984
	0.8	0.924	0.922	0.912	0.911	0.921	0.919	0.816	0.982	0.791	0.974
	0.9	0.644	0.640	0.664	0.659	0.649	0.644	0.616	0.771	0.558	0.707
	1.0	0.057	0.049	0.076	0.068	0.064	0.055	0.054	0.071	0.043	0.059
200	0.1	0.999	0.999	0.976	0.979	0.993	0.993	0.757	0.809	0.639	0.635
	0.2	0.999	0.999	0.980	0.982	0.995	0.995	0.778	0.879	0.666	0.773
	0.3	0.999	0.999	0.982	0.984	0.996	0.997	0.804	0.932	0.703	0.876
	0.4	1.000	1.000	0.982	0.983	0.996	0.997	0.825	0.958	0.736	0.930
	0.5	1.000	1.000	0.985	0.985	0.997	0.997	0.849	0.977	0.772	0.965
	0.6	0.999	0.999	0.987	0.988	0.998	0.998	0.868	0.991	0.811	0.987
	0.7	0.999	0.999	0.991	0.992	0.998	0.998	0.898	0.996	0.855	0.995
	0.8	0.997	0.998	0.991	0.992	0.996	0.997	0.920	0.999	0.894	0.998
	0.9	0.972	0.974	0.964	0.966	0.970	0.971	0.898	0.988	0.888	0.986
	1.0	0.053	0.050	0.059	0.057	0.054	0.051	0.046	0.057	0.042	0.052
400	0.1	0.999	0.999	0.987	0.989	0.997	0.997	0.792	0.880	0.674	0.623
	0.2	1.000	1.000	0.987	0.989	0.998	0.998	0.826	0.909	0.723	0.746
	0.3	1.000	1.000	0.989	0.989	0.999	0.999	0.844	0.891	0.743	0.756
	0.4	1.000	1.000	0.990	0.991	0.998	0.998	0.869	0.894	0.783	0.800
	0.5	1.000	1.000	0.994	0.994	0.999	1.000	0.891	0.925	0.825	0.872
	0.6	1.000	1.000	0.996	0.996	1.000	1.000	0.926	0.962	0.882	0.945
	0.7	1.000	1.000	0.998	0.998	1.000	1.000	0.952	0.991	0.922	0.989
	0.8	1.000	1.000	0.999	0.999	1.000	1.000	0.975	1.000	0.958	0.999
	0.9	1.000	1.000	0.999	0.999	1.000	1.000	0.986	1.000	0.982	1.000
	1.0	0.053	0.050	0.059	0.057	0.054	0.051	0.043	0.051	0.042	0.049

Table 3.17. Power of the tests for model with an intercept and trend, $\phi = 0$.

N	ρ	$\overline{BNM}_b$	$\overline{BEPO}_b$	$\overline{BNM}_a$	$\overline{BEPO}_a$	$\overline{BNM}_h$	$\overline{BEPO}_h$	ADF_{PQ}	ADF_S	M_{PQ}	M_S
50	0.1	0.950	0.926	0.833	0.786	0.904	0.867	0.726	0.987	0.604	0.972
	0.2	0.936	0.906	0.820	0.775	0.890	0.854	0.733	0.982	0.607	0.971
	0.3	0.919	0.887	0.807	0.759	0.880	0.845	0.735	0.981	0.607	0.953
	0.4	0.882	0.836	0.772	0.724	0.840	0.794	0.729	0.978	0.572	0.899
	0.5	0.821	0.752	0.734	0.668	0.792	0.726	0.712	0.966	0.471	0.748
	0.6	0.718	0.630	0.656	0.581	0.698	0.615	0.666	0.915	0.314	0.515
	0.7	0.542	0.447	0.547	0.464	0.551	0.461	0.537	0.748	0.159	0.289
	0.8	0.323	0.239	0.372	0.294	0.350	0.268	0.317	0.464	0.053	0.133
	0.9	0.149	0.105	0.208	0.162	0.176	0.131	0.138	0.221	0.017	0.062
	1.0	0.071	0.051	0.117	0.094	0.095	0.073	0.070	0.113	0.008	0.035
100	0.1	0.995	0.993	0.927	0.915	0.974	0.969	0.754	0.957	0.663	0.946
	0.2	0.993	0.992	0.929	0.916	0.977	0.973	0.765	0.971	0.687	0.965
	0.3	0.992	0.990	0.924	0.913	0.974	0.969	0.771	0.976	0.703	0.973
	0.4	0.990	0.985	0.920	0.907	0.970	0.963	0.779	0.984	0.727	0.983
	0.5	0.981	0.975	0.919	0.908	0.966	0.959	0.780	0.988	0.738	0.987
	0.6	0.958	0.948	0.910	0.896	0.945	0.935	0.781	0.991	0.739	0.989
	0.7	0.910	0.887	0.879	0.857	0.902	0.880	0.757	0.990	0.701	0.965
	0.8	0.733	0.685	0.740	0.699	0.736	0.692	0.636	0.879	0.497	0.722
	0.9	0.308	0.264	0.373	0.331	0.332	0.288	0.252	0.365	0.137	0.212
	1.0	0.062	0.050	0.099	0.084	0.076	0.062	0.046	0.068	0.019	0.033
200	0.1	0.998	0.998	0.960	0.954	0.988	0.987	0.796	0.877	0.680	0.839
	0.2	1.000	1.000	0.967	0.962	0.995	0.995	0.809	0.917	0.703	0.904
	0.3	1.000	1.000	0.963	0.959	0.994	0.993	0.820	0.960	0.727	0.961
	0.4	1.000	1.000	0.968	0.965	0.996	0.995	0.836	0.987	0.764	0.988
	0.5	1.000	1.000	0.967	0.963	0.994	0.993	0.840	0.997	0.785	0.996
	0.6	0.999	0.999	0.972	0.970	0.995	0.995	0.859	0.998	0.820	0.999
	0.7	0.995	0.994	0.981	0.979	0.993	0.992	0.863	0.999	0.844	0.999
	0.8	0.975	0.971	0.967	0.963	0.973	0.969	0.852	0.998	0.845	0.998
	0.9	0.745	0.723	0.757	0.738	0.749	0.727	0.671	0.856	0.608	0.783
	1.0	0.048	0.044	0.063	0.058	0.052	0.048	0.037	0.053	0.027	0.039
400	0.1	1.000	1.000	0.980	0.979	0.997	0.997	0.874	0.905	0.711	0.793
	0.2	1.000	1.000	0.978	0.978	0.998	0.998	0.885	0.961	0.741	0.898
	0.3	1.000	1.000	0.980	0.980	0.998	0.998	0.898	0.970	0.769	0.938
	0.4	1.000	1.000	0.981	0.980	0.999	0.999	0.921	0.977	0.818	0.961
	0.5	1.000	1.000	0.985	0.984	0.999	0.999	0.939	0.990	0.853	0.985
	0.6	1.000	1.000	0.989	0.988	0.999	0.999	0.953	0.997	0.895	0.996
	0.7	1.000	1.000	0.998	0.998	1.000	1.000	0.962	0.999	0.922	0.999
	0.8	1.000	1.000	0.998	0.997	0.999	0.999	0.966	1.000	0.949	1.000
	0.9	0.978	0.976	0.975	0.973	0.978	0.976	0.939	1.000	0.939	1.000
	1.0	0.051	0.049	0.056	0.054	0.052	0.050	0.037	0.049	0.034	0.042

Table 3.18. Relative frequencies of MA order chosen when only an intercept is included in the model.

		N = 50						N = 100					
ϕ		0	1	2	3	4	5	0	1	2	3	4	5
−0.8	BIC	0.686	0.249	0.043	0.019	0.004	0.001	0.223	0.731	0.033	0.008	0.003	0.002
	AIC	0.394	0.278	0.110	0.087	0.070	0.062	0.094	0.603	0.116	0.072	0.059	0.057
	HQIC	0.527	0.291	0.082	0.054	0.028	0.019	0.152	0.713	0.076	0.029	0.019	0.013
−0.7	BIC	0.497	0.416	0.057	0.019	0.011	0.001	0.093	0.863	0.031	0.008	0.004	0.001
	AIC	0.271	0.366	0.117	0.085	0.092	0.071	0.031	0.673	0.112	0.072	0.056	0.055
	HQIC	0.378	0.406	0.099	0.055	0.044	0.020	0.055	0.814	0.071	0.030	0.017	0.012
−0.6	BIC	0.423	0.485	0.058	0.022	0.008	0.005	0.050	0.904	0.034	0.009	0.003	0.001
	AIC	0.199	0.455	0.101	0.088	0.086	0.072	0.014	0.681	0.121	0.071	0.056	0.056
	HQIC	0.296	0.501	0.086	0.055	0.042	0.021	0.027	0.834	0.076	0.033	0.018	0.012
−0.5	BIC	0.451	0.458	0.058	0.028	0.005	0.002	0.070	0.889	0.031	0.007	0.002	0.001
	AIC	0.207	0.432	0.112	0.098	0.083	0.069	0.014	0.700	0.116	0.069	0.048	0.053
	HQIC	0.311	0.485	0.090	0.060	0.035	0.021	0.032	0.843	0.072	0.029	0.013	0.012
−0.4	BIC	0.591	0.333	0.045	0.021	0.008	0.003	0.164	0.792	0.035	0.007	0.001	0.001
	AIC	0.298	0.353	0.109	0.091	0.082	0.069	0.039	0.667	0.123	0.070	0.050	0.051
	HQIC	0.438	0.384	0.086	0.047	0.033	0.015	0.082	0.786	0.077	0.032	0.014	0.009
−0.3	BIC	0.736	0.212	0.035	0.011	0.005	0.002	0.398	0.567	0.027	0.005	0.002	0.001
	AIC	0.410	0.283	0.110	0.071	0.058	0.070	0.132	0.592	0.108	0.064	0.053	0.051
	HQIC	0.557	0.283	0.071	0.040	0.028	0.023	0.240	0.647	0.064	0.023	0.015	0.010
−0.2	BIC	0.869	0.090	0.024	0.015	0.002	0.001	0.726	0.246	0.020	0.005	0.002	0.001
	AIC	0.561	0.176	0.075	0.069	0.062	0.058	0.365	0.381	0.102	0.062	0.044	0.046
	HQIC	0.717	0.150	0.051	0.042	0.024	0.018	0.546	0.353	0.057	0.023	0.012	0.009
−0.1	BIC	0.921	0.045	0.019	0.011	0.004	0.001	0.928	0.060	0.009	0.002	0.001	0.000
	AIC	0.666	0.091	0.063	0.061	0.065	0.056	0.641	0.166	0.068	0.047	0.039	0.039
	HQIC	0.814	0.072	0.045	0.031	0.026	0.013	0.821	0.115	0.033	0.014	0.010	0.008
0	BIC	0.920	0.055	0.014	0.007	0.005	0.001	0.959	0.031	0.006	0.002	0.001	0.000
	AIC	0.645	0.115	0.069	0.062	0.062	0.049	0.706	0.116	0.062	0.043	0.035	0.038
	HQIC	0.794	0.091	0.043	0.034	0.026	0.012	0.875	0.073	0.027	0.012	0.009	0.005
0.1	BIC	0.822	0.138	0.027	0.007	0.007	0.001	0.824	0.156	0.014	0.004	0.001	0.001
	AIC	0.525	0.208	0.086	0.057	0.063	0.063	0.492	0.289	0.085	0.053	0.042	0.040
	HQIC	0.672	0.188	0.058	0.035	0.034	0.014	0.676	0.247	0.044	0.019	0.009	0.006
0.2	BIC	0.624	0.317	0.037	0.013	0.009	0.001	0.500	0.468	0.025	0.005	0.002	0.001
	AIC	0.328	0.381	0.102	0.074	0.064	0.053	0.188	0.549	0.104	0.062	0.049	0.048
	HQIC	0.460	0.383	0.072	0.042	0.031	0.014	0.320	0.574	0.060	0.024	0.012	0.010
0.3	BIC	0.378	0.547	0.045	0.018	0.011	0.002	0.163	0.794	0.034	0.007	0.002	0.001
	AIC	0.140	0.517	0.114	0.087	0.072	0.072	0.034	0.672	0.121	0.070	0.054	0.050
	HQIC	0.241	0.579	0.086	0.044	0.034	0.018	0.076	0.795	0.079	0.026	0.016	0.010
0.4	BIC	0.167	0.754	0.047	0.021	0.011	0.002	0.025	0.932	0.033	0.007	0.002	0.002
	AIC	0.039	0.612	0.120	0.081	0.072	0.077	0.003	0.709	0.122	0.066	0.052	0.048
	HQIC	0.085	0.713	0.094	0.049	0.034	0.027	0.009	0.863	0.076	0.028	0.014	0.010
0.5	BIC	0.038	0.877	0.058	0.016	0.010	0.003	0.002	0.952	0.034	0.008	0.003	0.001
	AIC	0.009	0.650	0.127	0.084	0.074	0.058	0.000	0.710	0.117	0.074	0.053	0.046
	HQIC	0.018	0.777	0.103	0.049	0.037	0.017	0.000	0.866	0.074	0.031	0.018	0.010
0.6	BIC	0.006	0.912	0.055	0.019	0.007	0.003	0.000	0.954	0.034	0.007	0.003	0.002
	AIC	0.001	0.664	0.123	0.077	0.073	0.064	0.000	0.702	0.126	0.074	0.050	0.049
	HQIC	0.002	0.804	0.093	0.045	0.036	0.022	0.000	0.865	0.079	0.030	0.015	0.011
0.7	BIC	0.001	0.916	0.055	0.016	0.011	0.003	0.000	0.956	0.034	0.007	0.002	0.001
	AIC	0.000	0.682	0.117	0.068	0.072	0.062	0.000	0.712	0.126	0.067	0.046	0.048
	HQIC	0.000	0.816	0.094	0.039	0.035	0.017	0.000	0.873	0.077	0.028	0.013	0.009
0.8	BIC	0.001	0.912	0.059	0.021	0.007	0.002	0.000	0.957	0.034	0.006	0.002	0.001
	AIC	0.000	0.659	0.132	0.066	0.080	0.064	0.000	0.713	0.121	0.067	0.053	0.046
	HQIC	0.001	0.801	0.095	0.043	0.037	0.024	0.000	0.870	0.075	0.029	0.017	0.009

Table 3.19. Relative frequencies of MA order chosen when only an intercept is included in the model.

		$N = 200$						$N = 400$					
ϕ		0	1	2	3	4	5	0	1	2	3	4	5
−0.8	BIC	0.014	0.958	0.023	0.004	0.001	0.000	0.000	0.983	0.015	0.001	0.000	0.000
	AIC	0.003	0.721	0.121	0.069	0.045	0.042	0.000	0.730	0.123	0.065	0.048	0.035
	HQIC	0.007	0.893	0.062	0.023	0.009	0.006	0.000	0.918	0.056	0.017	0.006	0.002
−0.7	BIC	0.002	0.971	0.023	0.003	0.001	0.000	0.000	0.984	0.015	0.001	0.000	0.000
	AIC	0.001	0.722	0.123	0.068	0.047	0.039	0.000	0.744	0.117	0.064	0.040	0.036
	HQIC	0.001	0.898	0.066	0.022	0.010	0.004	0.000	0.924	0.053	0.016	0.005	0.002
−0.6	BIC	0.001	0.972	0.024	0.003	0.001	0.000	0.000	0.984	0.014	0.002	0.000	0.000
	AIC	0.000	0.724	0.122	0.067	0.047	0.040	0.000	0.734	0.125	0.068	0.041	0.033
	HQIC	0.000	0.897	0.067	0.022	0.010	0.004	0.000	0.917	0.056	0.018	0.006	0.004
−0.5	BIC	0.001	0.973	0.024	0.002	0.000	0.000	0.000	0.984	0.015	0.001	0.000	0.000
	AIC	0.000	0.734	0.119	0.064	0.045	0.038	0.000	0.737	0.121	0.064	0.041	0.037
	HQIC	0.000	0.906	0.062	0.019	0.009	0.003	0.000	0.923	0.055	0.015	0.005	0.002
−0.4	BIC	0.003	0.971	0.022	0.004	0.001	0.000	0.000	0.985	0.013	0.001	0.000	0.000
	AIC	0.000	0.731	0.119	0.065	0.045	0.039	0.000	0.741	0.124	0.062	0.041	0.032
	HQIC	0.001	0.901	0.063	0.020	0.010	0.005	0.000	0.925	0.055	0.014	0.004	0.003
−0.3	BIC	0.065	0.909	0.023	0.003	0.000	0.000	0.001	0.982	0.016	0.002	0.000	0.000
	AIC	0.007	0.725	0.123	0.066	0.043	0.037	0.000	0.739	0.122	0.063	0.040	0.036
	HQIC	0.023	0.881	0.064	0.020	0.008	0.004	0.000	0.917	0.059	0.016	0.006	0.003
−0.2	BIC	0.420	0.557	0.019	0.003	0.000	0.000	0.092	0.892	0.015	0.001	0.000	0.000
	AIC	0.111	0.628	0.115	0.064	0.046	0.036	0.007	0.732	0.115	0.064	0.045	0.037
	HQIC	0.237	0.673	0.059	0.018	0.008	0.005	0.029	0.894	0.054	0.015	0.006	0.003
−0.1	BIC	0.868	0.123	0.008	0.001	0.000	0.000	0.729	0.263	0.007	0.001	0.000	0.000
	AIC	0.503	0.290	0.085	0.051	0.040	0.031	0.286	0.486	0.101	0.057	0.040	0.029
	HQIC	0.716	0.232	0.032	0.013	0.006	0.002	0.512	0.433	0.038	0.011	0.005	0.001
0	BIC	0.975	0.022	0.003	0.001	0.000	0.000	0.983	0.016	0.002	0.000	0.000	0.000
	AIC	0.715	0.124	0.062	0.039	0.032	0.028	0.717	0.123	0.064	0.042	0.030	0.023
	HQIC	0.901	0.066	0.020	0.009	0.003	0.002	0.915	0.060	0.015	0.007	0.003	0.001
0.1	BIC	0.772	0.216	0.010	0.002	0.000	0.000	0.636	0.353	0.009	0.001	0.000	0.000
	AIC	0.369	0.408	0.093	0.055	0.039	0.036	0.204	0.560	0.107	0.056	0.043	0.031
	HQIC	0.583	0.354	0.041	0.013	0.006	0.004	0.406	0.532	0.044	0.012	0.004	0.002
0.2	BIC	0.258	0.716	0.022	0.003	0.000	0.000	0.051	0.935	0.012	0.001	0.000	0.000
	AIC	0.050	0.687	0.121	0.063	0.042	0.037	0.003	0.739	0.118	0.065	0.041	0.036
	HQIC	0.122	0.788	0.058	0.020	0.008	0.005	0.013	0.911	0.052	0.017	0.005	0.002
0.3	BIC	0.022	0.957	0.018	0.003	0.000	0.000	0.000	0.985	0.013	0.002	0.000	0.000
	AIC	0.001	0.731	0.119	0.066	0.046	0.036	0.000	0.745	0.119	0.063	0.040	0.034
	HQIC	0.005	0.900	0.063	0.018	0.011	0.004	0.000	0.921	0.056	0.016	0.005	0.002
0.4	BIC	0.000	0.974	0.023	0.003	0.000	0.000	0.000	0.984	0.014	0.002	0.000	0.000
	AIC	0.000	0.725	0.121	0.067	0.046	0.040	0.000	0.730	0.120	0.069	0.045	0.035
	HQIC	0.000	0.901	0.063	0.022	0.009	0.004	0.000	0.923	0.052	0.018	0.004	0.003
0.5	BIC	0.000	0.974	0.022	0.003	0.001	0.000	0.000	0.986	0.013	0.001	0.000	0.000
	AIC	0.000	0.727	0.124	0.064	0.047	0.038	0.000	0.741	0.116	0.064	0.044	0.036
	HQIC	0.000	0.903	0.064	0.022	0.008	0.004	0.000	0.927	0.051	0.013	0.006	0.003
0.6	BIC	0.000	0.972	0.024	0.003	0.001	0.000	0.000	0.984	0.015	0.001	0.000	0.000
	AIC	0.000	0.730	0.120	0.066	0.044	0.040	0.000	0.743	0.117	0.065	0.042	0.033
	HQIC	0.000	0.902	0.068	0.019	0.008	0.004	0.000	0.926	0.052	0.014	0.006	0.002
0.7	BIC	0.000	0.976	0.021	0.003	0.001	0.000	0.000	0.981	0.017	0.002	0.000	0.000
	AIC	0.000	0.727	0.121	0.066	0.047	0.039	0.000	0.739	0.119	0.066	0.041	0.035
	HQIC	0.000	0.904	0.065	0.020	0.009	0.003	0.000	0.920	0.058	0.015	0.005	0.002
0.8	BIC	0.000	0.973	0.023	0.003	0.001	0.000	0.000	0.983	0.016	0.001	0.000	0.000
	AIC	0.000	0.736	0.117	0.065	0.044	0.038	0.000	0.737	0.124	0.062	0.043	0.034
	HQIC	0.000	0.906	0.065	0.019	0.007	0.003	0.000	0.917	0.060	0.014	0.006	0.003

Table 3.20. Relative frequencies of MA order chosen when only an intercept is included in the model.

		$N = 50$						$N = 100$					
ρ		0	1	2	3	4	5	0	1	2	3	4	5
0.1	BIC	0.910	0.054	0.022	0.010	0.003	0.001	0.939	0.044	0.014	0.002	0.001	0.000
	AIC	0.605	0.126	0.095	0.070	0.058	0.045	0.615	0.135	0.094	0.061	0.047	0.048
	$HQIC$	0.777	0.099	0.058	0.036	0.020	0.010	0.824	0.092	0.047	0.021	0.010	0.007
0.2	BIC	0.915	0.049	0.023	0.009	0.003	0.001	0.946	0.040	0.012	0.001	0.001	0.000
	AIC	0.623	0.109	0.092	0.069	0.058	0.049	0.641	0.128	0.086	0.057	0.047	0.041
	$HQIC$	0.794	0.082	0.055	0.033	0.022	0.013	0.843	0.086	0.040	0.016	0.009	0.007
0.3	BIC	0.914	0.053	0.023	0.007	0.003	0.000	0.951	0.037	0.008	0.003	0.001	0.000
	AIC	0.626	0.112	0.087	0.063	0.064	0.049	0.663	0.113	0.076	0.058	0.048	0.042
	$HQIC$	0.788	0.092	0.056	0.032	0.020	0.012	0.855	0.073	0.035	0.020	0.011	0.006
0.4	BIC	0.908	0.059	0.021	0.008	0.003	0.001	0.948	0.039	0.009	0.002	0.000	0.000
	AIC	0.635	0.116	0.082	0.058	0.058	0.051	0.671	0.112	0.072	0.058	0.041	0.046
	$HQIC$	0.784	0.096	0.054	0.030	0.023	0.012	0.855	0.077	0.032	0.021	0.007	0.009
0.5	BIC	0.912	0.055	0.022	0.006	0.003	0.001	0.954	0.033	0.010	0.003	0.001	0.000
	AIC	0.643	0.115	0.074	0.058	0.061	0.049	0.681	0.112	0.075	0.050	0.046	0.036
	$HQIC$	0.791	0.094	0.054	0.027	0.023	0.011	0.860	0.074	0.034	0.016	0.009	0.006
0.6	BIC	0.909	0.055	0.025	0.006	0.004	0.001	0.954	0.034	0.008	0.002	0.001	0.000
	AIC	0.657	0.114	0.073	0.051	0.059	0.045	0.686	0.106	0.072	0.049	0.048	0.040
	$HQIC$	0.802	0.087	0.050	0.026	0.023	0.013	0.861	0.071	0.034	0.014	0.012	0.007
0.7	BIC	0.920	0.049	0.022	0.006	0.002	0.001	0.955	0.031	0.009	0.002	0.003	0.000
	AIC	0.658	0.110	0.077	0.062	0.053	0.041	0.690	0.108	0.066	0.049	0.045	0.041
	$HQIC$	0.808	0.082	0.051	0.027	0.022	0.011	0.868	0.072	0.031	0.013	0.010	0.007
0.8	BIC	0.913	0.052	0.022	0.007	0.004	0.001	0.959	0.031	0.006	0.003	0.001	0.000
	AIC	0.647	0.112	0.075	0.059	0.066	0.041	0.696	0.116	0.057	0.048	0.043	0.039
	$HQIC$	0.794	0.091	0.049	0.029	0.027	0.009	0.868	0.072	0.027	0.016	0.011	0.006
0.9	BIC	0.915	0.051	0.019	0.010	0.004	0.001	0.958	0.032	0.008	0.002	0.001	0.000
	AIC	0.649	0.118	0.066	0.058	0.058	0.051	0.713	0.113	0.060	0.040	0.036	0.038
	$HQIC$	0.807	0.085	0.042	0.029	0.024	0.012	0.874	0.072	0.030	0.010	0.007	0.006
1	BIC	0.913	0.057	0.018	0.007	0.004	0.000	0.956	0.035	0.005	0.002	0.001	0.001
	AIC	0.649	0.115	0.069	0.063	0.059	0.046	0.708	0.114	0.063	0.048	0.030	0.037
	$HQIC$	0.803	0.094	0.039	0.029	0.024	0.011	0.871	0.075	0.026	0.016	0.006	0.006

		$N = 200$						$N = 400$					
ρ		0	1	2	3	4	5	0	1	2	3	4	5
0.1	BIC	0.956	0.034	0.009	0.000	0.000	0.000	0.976	0.020	0.004	0.000	0.000	0.000
	AIC	0.630	0.140	0.094	0.060	0.043	0.032	0.637	0.142	0.096	0.056	0.040	0.029
	$HQIC$	0.855	0.090	0.036	0.011	0.006	0.002	0.890	0.071	0.028	0.007	0.003	0.001
0.2	BIC	0.963	0.030	0.006	0.001	0.000	0.000	0.980	0.017	0.003	0.000	0.000	0.000
	AIC	0.655	0.128	0.087	0.054	0.039	0.037	0.679	0.108	0.084	0.054	0.040	0.034
	$HQIC$	0.880	0.073	0.028	0.013	0.004	0.002	0.906	0.056	0.025	0.009	0.003	0.001
0.3	BIC	0.974	0.019	0.005	0.002	0.000	0.000	0.980	0.016	0.004	0.000	0.000	0.000
	AIC	0.677	0.110	0.073	0.059	0.044	0.037	0.696	0.109	0.070	0.055	0.037	0.033
	$HQIC$	0.890	0.062	0.029	0.012	0.004	0.002	0.912	0.056	0.021	0.008	0.002	0.001
0.4	BIC	0.973	0.019	0.005	0.002	0.000	0.000	0.981	0.017	0.002	0.000	0.000	0.000
	AIC	0.688	0.106	0.070	0.056	0.044	0.036	0.699	0.106	0.068	0.054	0.037	0.035
	$HQIC$	0.892	0.056	0.028	0.015	0.005	0.003	0.912	0.055	0.018	0.010	0.004	0.001
0.5	BIC	0.971	0.023	0.006	0.001	0.000	0.000	0.982	0.017	0.001	0.000	0.000	0.000
	AIC	0.690	0.117	0.062	0.057	0.038	0.035	0.695	0.118	0.062	0.047	0.041	0.036
	$HQIC$	0.894	0.065	0.021	0.012	0.005	0.003	0.912	0.058	0.020	0.005	0.002	0.003
0.6	BIC	0.980	0.017	0.003	0.000	0.000	0.000	0.984	0.013	0.002	0.000	0.000	0.000
	AIC	0.714	0.107	0.061	0.045	0.037	0.035	0.713	0.123	0.063	0.041	0.030	0.030
	$HQIC$	0.902	0.061	0.019	0.011	0.004	0.003	0.919	0.056	0.016	0.007	0.001	0.002
0.7	BIC	0.973	0.024	0.002	0.001	0.000	0.000	0.985	0.015	0.001	0.000	0.000	0.000
	AIC	0.715	0.118	0.063	0.041	0.033	0.030	0.735	0.109	0.061	0.039	0.033	0.023
	$HQIC$	0.901	0.063	0.019	0.009	0.005	0.003	0.925	0.052	0.013	0.006	0.002	0.001
0.8	BIC	0.972	0.023	0.004	0.001	0.000	0.000	0.979	0.020	0.002	0.000	0.000	0.000
	AIC	0.721	0.108	0.064	0.043	0.036	0.028	0.719	0.122	0.060	0.040	0.036	0.023
	$HQIC$	0.905	0.059	0.019	0.010	0.005	0.002	0.911	0.063	0.017	0.005	0.003	0.001
0.9	BIC	0.978	0.020	0.002	0.000	0.000	0.000	0.985	0.013	0.002	0.000	0.000	0.000
	AIC	0.722	0.118	0.061	0.043	0.029	0.026	0.731	0.115	0.061	0.041	0.031	0.021
	$HQIC$	0.909	0.063	0.018	0.006	0.003	0.001	0.927	0.050	0.014	0.007	0.002	0.000
1	BIC	0.972	0.025	0.002	0.001	0.000	0.000	0.984	0.015	0.001	0.001	0.000	0.000
	AIC	0.713	0.126	0.057	0.044	0.032	0.028	0.732	0.112	0.060	0.042	0.028	0.026
	$HQIC$	0.898	0.069	0.018	0.008	0.005	0.003	0.926	0.051	0.015	0.005	0.002	0.001

Table 3.21. Relative frequencies of MA order chosen when an intercept and a trend are included in the model.

		$N = 50$						$N = 100$					
ρ		0	1	2	3	4	5	0	1	2	3	4	5
0.1	BIC	0.870	0.077	0.038	0.011	0.003	0.001	0.920	0.055	0.020	0.004	0.002	0.000
	AIC	0.515	0.149	0.130	0.082	0.061	0.064	0.555	0.151	0.125	0.072	0.050	0.048
	$HQIC$	0.702	0.126	0.087	0.045	0.022	0.018	0.775	0.108	0.070	0.025	0.012	0.010
0.2	BIC	0.877	0.060	0.039	0.018	0.004	0.001	0.925	0.044	0.024	0.005	0.001	0.001
	AIC	0.539	0.113	0.132	0.090	0.067	0.059	0.586	0.113	0.123	0.078	0.048	0.052
	$HQIC$	0.719	0.097	0.092	0.050	0.026	0.016	0.798	0.086	0.069	0.026	0.012	0.009
0.3	BIC	0.894	0.052	0.034	0.015	0.005	0.001	0.941	0.035	0.018	0.004	0.001	0.001
	AIC	0.566	0.097	0.121	0.094	0.065	0.056	0.611	0.098	0.098	0.085	0.057	0.052
	$HQIC$	0.740	0.087	0.081	0.050	0.023	0.018	0.827	0.070	0.052	0.029	0.012	0.010
0.4	BIC	0.897	0.053	0.031	0.013	0.004	0.002	0.947	0.035	0.012	0.005	0.001	0.000
	AIC	0.588	0.097	0.106	0.084	0.066	0.059	0.633	0.102	0.075	0.079	0.061	0.050
	$HQIC$	0.756	0.082	0.074	0.044	0.027	0.018	0.835	0.072	0.042	0.029	0.014	0.008
0.5	BIC	0.908	0.052	0.025	0.010	0.003	0.001	0.948	0.036	0.011	0.003	0.002	0.000
	AIC	0.600	0.100	0.093	0.079	0.071	0.056	0.640	0.106	0.070	0.067	0.065	0.052
	$HQIC$	0.774	0.083	0.067	0.036	0.025	0.015	0.844	0.072	0.035	0.024	0.017	0.009
0.6	BIC	0.905	0.057	0.025	0.008	0.004	0.001	0.958	0.031	0.007	0.002	0.002	0.000
	AIC	0.613	0.109	0.085	0.073	0.063	0.058	0.668	0.105	0.058	0.059	0.057	0.054
	$HQIC$	0.780	0.094	0.056	0.034	0.025	0.012	0.861	0.070	0.028	0.019	0.015	0.008
0.7	BIC	0.906	0.054	0.027	0.008	0.004	0.001	0.953	0.036	0.008	0.002	0.002	0.000
	AIC	0.611	0.115	0.079	0.070	0.074	0.051	0.674	0.114	0.063	0.050	0.048	0.050
	$HQIC$	0.782	0.092	0.054	0.031	0.027	0.014	0.852	0.080	0.033	0.016	0.012	0.007
0.8	BIC	0.907	0.057	0.021	0.012	0.003	0.001	0.952	0.035	0.008	0.003	0.001	0.000
	AIC	0.627	0.110	0.069	0.068	0.076	0.050	0.671	0.119	0.066	0.049	0.046	0.049
	$HQIC$	0.786	0.089	0.049	0.037	0.028	0.011	0.854	0.077	0.028	0.019	0.013	0.009
0.9	BIC	0.896	0.063	0.026	0.009	0.004	0.001	0.948	0.040	0.009	0.002	0.001	0.001
	AIC	0.610	0.116	0.081	0.070	0.075	0.049	0.679	0.120	0.073	0.047	0.038	0.043
	$HQIC$	0.775	0.098	0.054	0.036	0.028	0.008	0.858	0.082	0.031	0.014	0.010	0.006
1	BIC	0.899	0.059	0.026	0.012	0.003	0.001	0.948	0.040	0.009	0.001	0.001	0.001
	AIC	0.597	0.111	0.084	0.080	0.073	0.055	0.678	0.124	0.069	0.045	0.041	0.044
	$HQIC$	0.763	0.096	0.056	0.045	0.029	0.011	0.853	0.082	0.032	0.015	0.010	0.008

		$N = 200$						$N = 400$					
ρ		0	1	2	3	4	5	0	1	2	3	4	5
0.1	BIC	0.952	0.036	0.011	0.002	0.001	0.000	0.971	0.023	0.006	0.000	0.000	0.000
	AIC	0.579	0.140	0.124	0.070	0.047	0.041	0.613	0.130	0.114	0.063	0.043	0.038
	$HQIC$	0.832	0.086	0.053	0.018	0.007	0.004	0.871	0.066	0.044	0.012	0.004	0.002
0.2	BIC	0.958	0.030	0.010	0.002	0.000	0.000	0.978	0.016	0.005	0.001	0.000	0.000
	AIC	0.618	0.108	0.113	0.072	0.050	0.041	0.643	0.103	0.098	0.071	0.050	0.035
	$HQIC$	0.856	0.067	0.047	0.019	0.007	0.004	0.888	0.056	0.035	0.014	0.005	0.002
0.3	BIC	0.963	0.027	0.008	0.002	0.000	0.000	0.982	0.014	0.004	0.001	0.000	0.000
	AIC	0.636	0.098	0.089	0.080	0.053	0.044	0.657	0.102	0.070	0.077	0.052	0.042
	$HQIC$	0.868	0.061	0.038	0.023	0.007	0.004	0.900	0.051	0.024	0.018	0.005	0.001
0.4	BIC	0.968	0.024	0.006	0.002	0.001	0.000	0.982	0.016	0.002	0.001	0.000	0.000
	AIC	0.659	0.101	0.066	0.073	0.058	0.044	0.674	0.109	0.060	0.062	0.052	0.044
	$HQIC$	0.880	0.060	0.029	0.019	0.009	0.004	0.906	0.052	0.018	0.015	0.007	0.003
0.5	BIC	0.972	0.022	0.004	0.002	0.001	0.000	0.982	0.016	0.002	0.000	0.000	0.000
	AIC	0.679	0.101	0.058	0.055	0.058	0.048	0.698	0.113	0.058	0.040	0.049	0.043
	$HQIC$	0.887	0.062	0.022	0.015	0.010	0.004	0.916	0.052	0.014	0.009	0.006	0.003
0.6	BIC	0.973	0.023	0.004	0.001	0.000	0.000	0.983	0.015	0.002	0.000	0.000	0.000
	AIC	0.696	0.108	0.059	0.043	0.048	0.046	0.710	0.115	0.062	0.036	0.033	0.044
	$HQIC$	0.898	0.064	0.020	0.009	0.006	0.004	0.918	0.055	0.016	0.006	0.004	0.002
0.7	BIC	0.973	0.022	0.004	0.000	0.000	0.000	0.982	0.016	0.002	0.000	0.000	0.000
	AIC	0.706	0.112	0.064	0.046	0.035	0.037	0.728	0.114	0.059	0.039	0.033	0.028
	$HQIC$	0.898	0.063	0.022	0.009	0.006	0.002	0.919	0.056	0.016	0.005	0.002	0.002
0.8	BIC	0.974	0.021	0.004	0.001	0.000	0.000	0.986	0.013	0.001	0.000	0.000	0.000
	AIC	0.712	0.118	0.063	0.045	0.032	0.030	0.725	0.118	0.063	0.039	0.031	0.023
	$HQIC$	0.899	0.065	0.021	0.010	0.004	0.002	0.923	0.053	0.016	0.006	0.003	0.001
0.9	BIC	0.973	0.023	0.004	0.001	0.000	0.000	0.982	0.016	0.002	0.000	0.000	0.000
	AIC	0.724	0.112	0.064	0.041	0.032	0.027	0.728	0.119	0.062	0.040	0.028	0.024
	$HQIC$	0.902	0.064	0.020	0.009	0.003	0.002	0.918	0.057	0.016	0.006	0.002	0.001
1	BIC	0.970	0.026	0.003	0.001	0.000	0.000	0.982	0.016	0.002	0.001	0.000	0.000
	AIC	0.702	0.117	0.070	0.044	0.035	0.032	0.722	0.120	0.062	0.042	0.031	0.025
	$HQIC$	0.889	0.071	0.024	0.010	0.004	0.002	0.918	0.058	0.016	0.007	0.001	0.001

Figure 3.1. Relative frequencies of information criteria for $\rho = 1$, $N = 100$, model with an intercept and a trend.

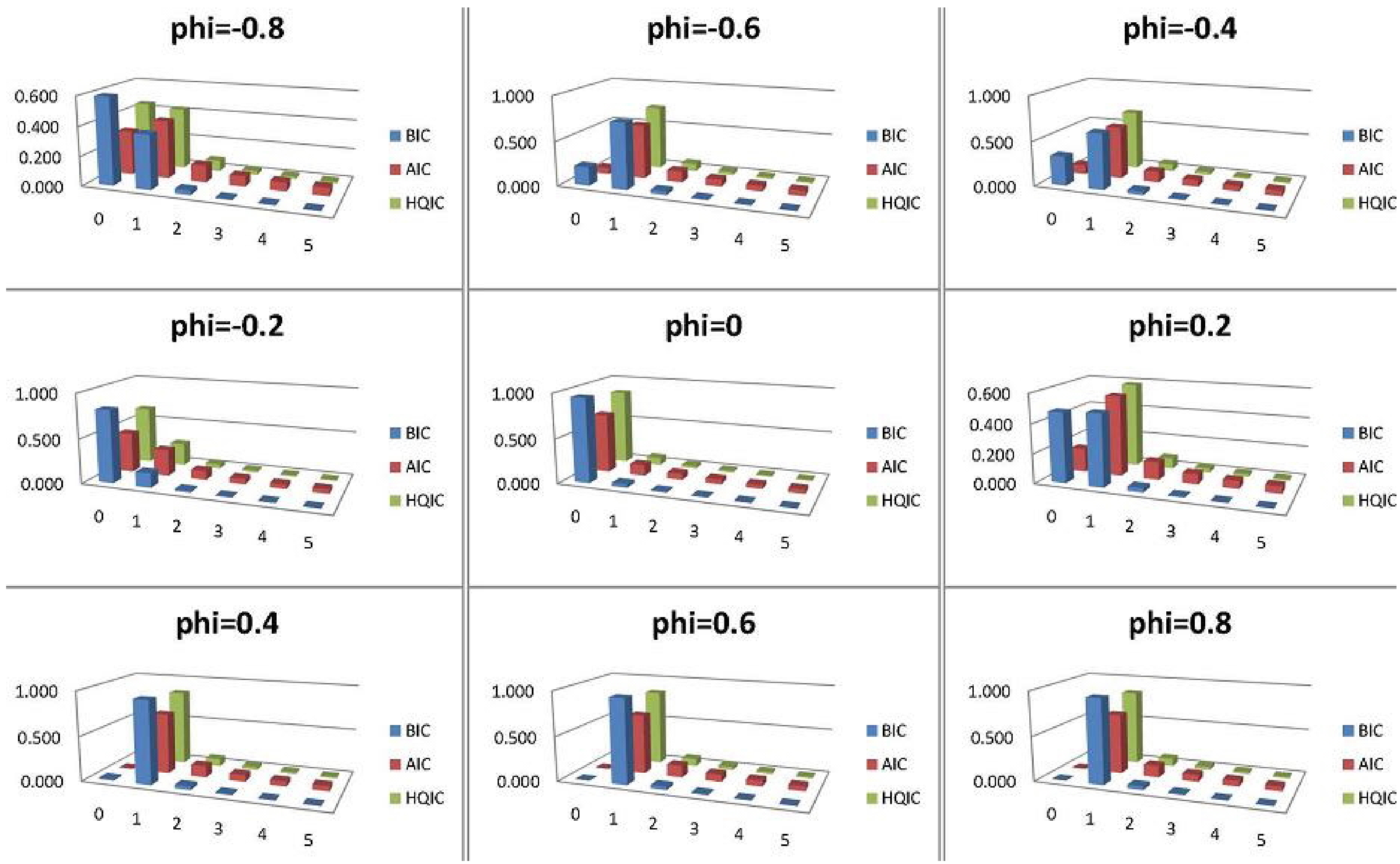


Figure 3.2. Empirical size of tests for a model with an intercept only.

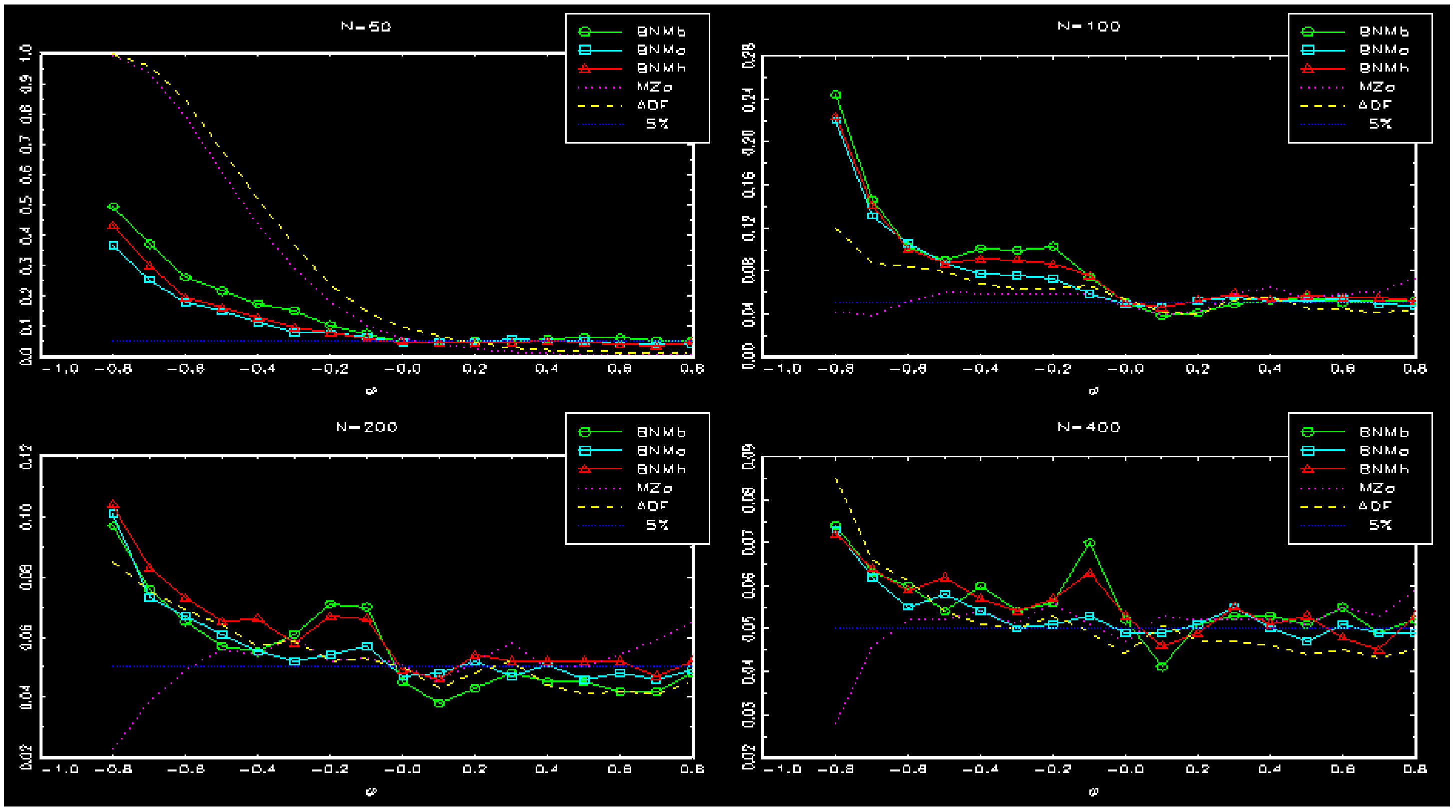


Figure 3.3. Empirical size for a model with an intercept and a trend.

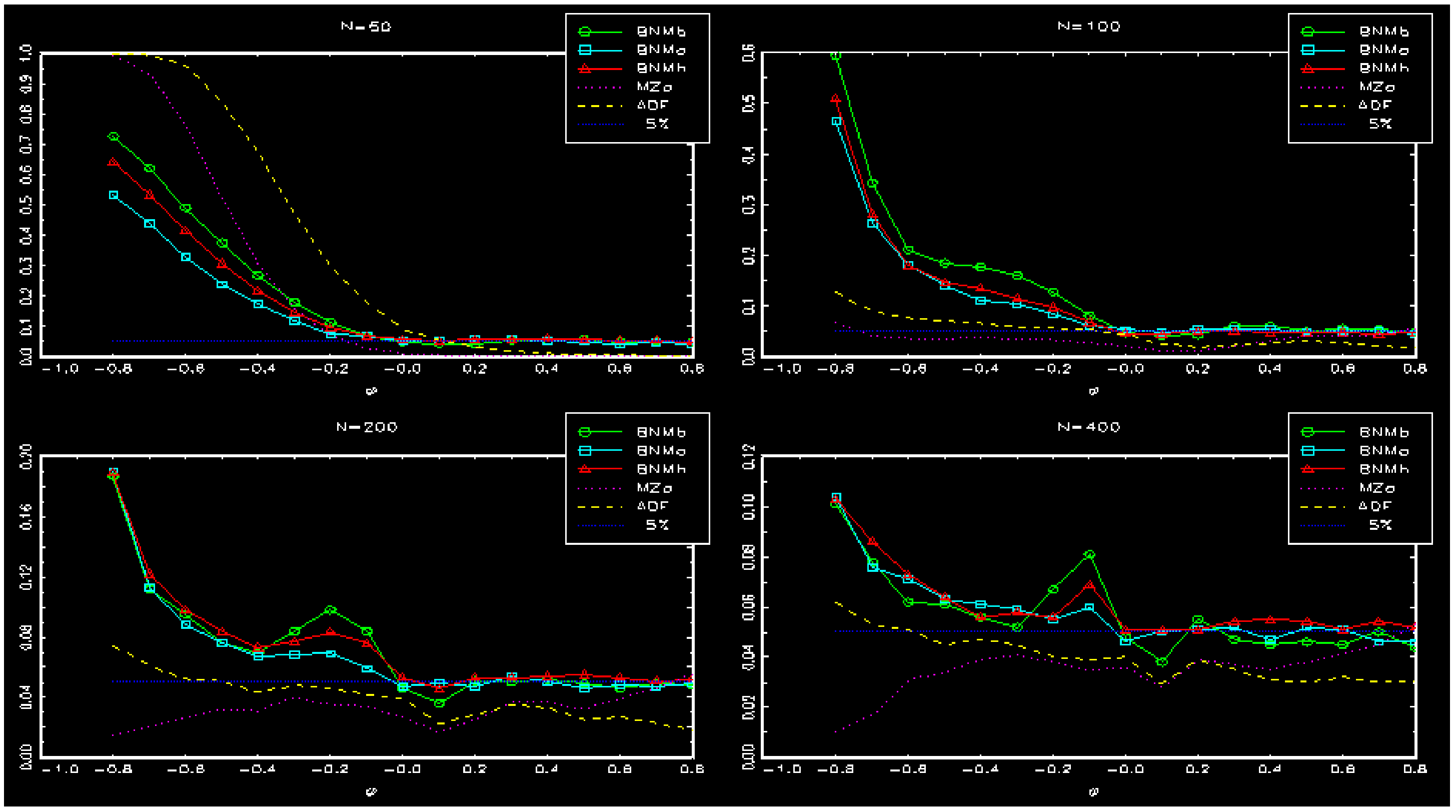


Figure 3.4. Power for model with an intercept only, $\phi = 0$.

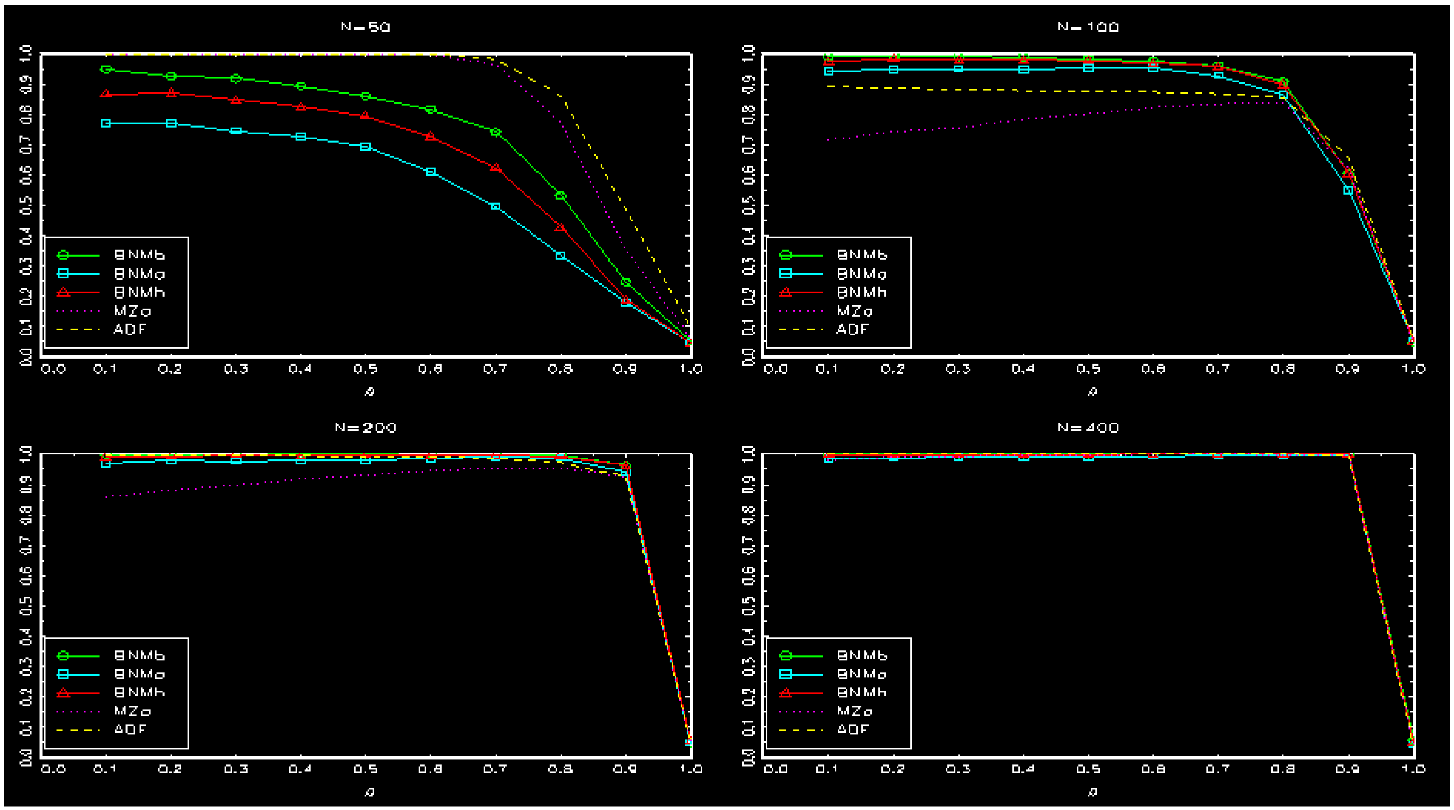


Figure 3.5. Power for model with an intercept and a trend, $\phi = 0$.

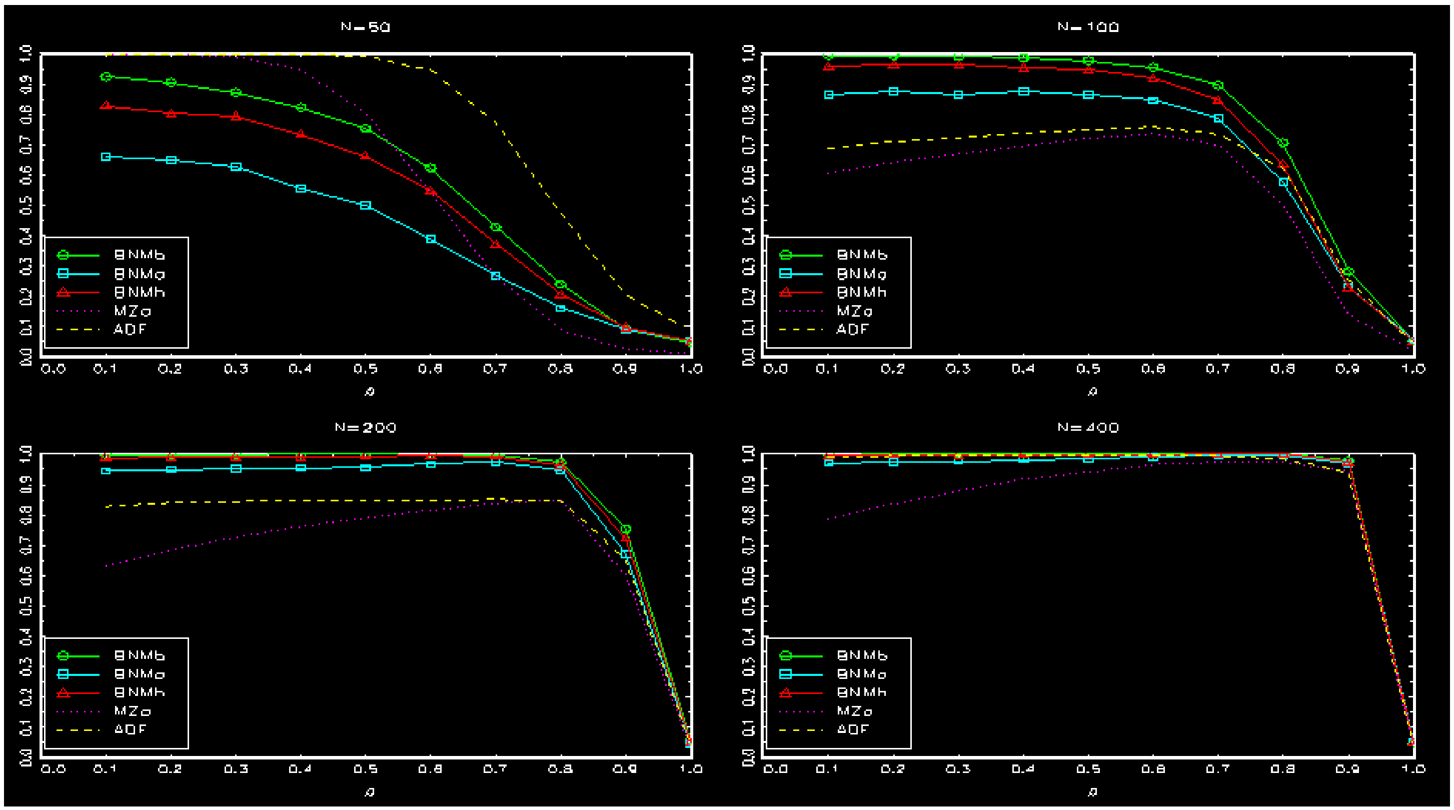


Figure 3.6. Empirical size of tests for a model with an intercept only. Asymptotic critical values used.

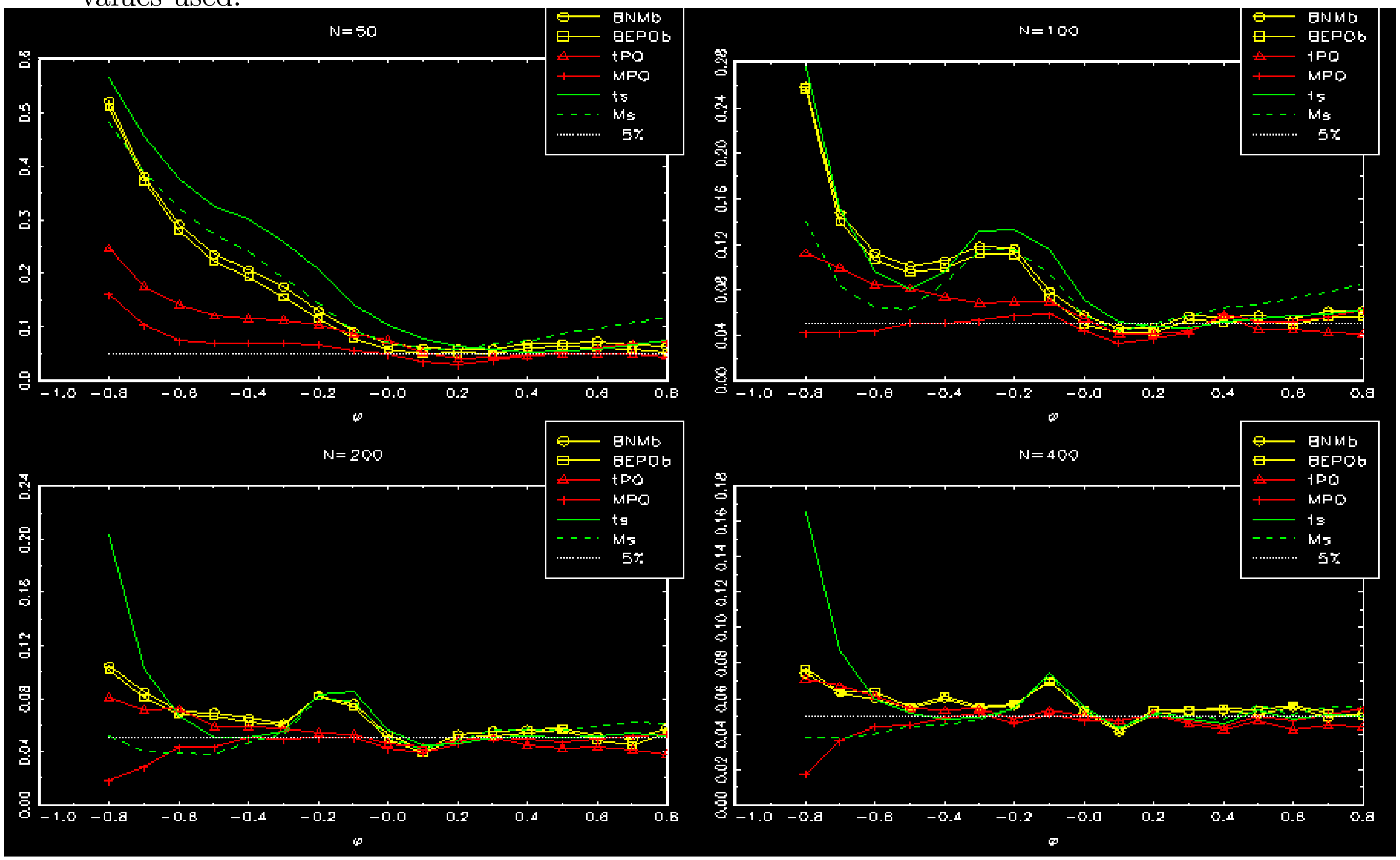


Figure 3.7. Empirical size of tests for a model with an intercept and a trend. Asymptotic critical values used.

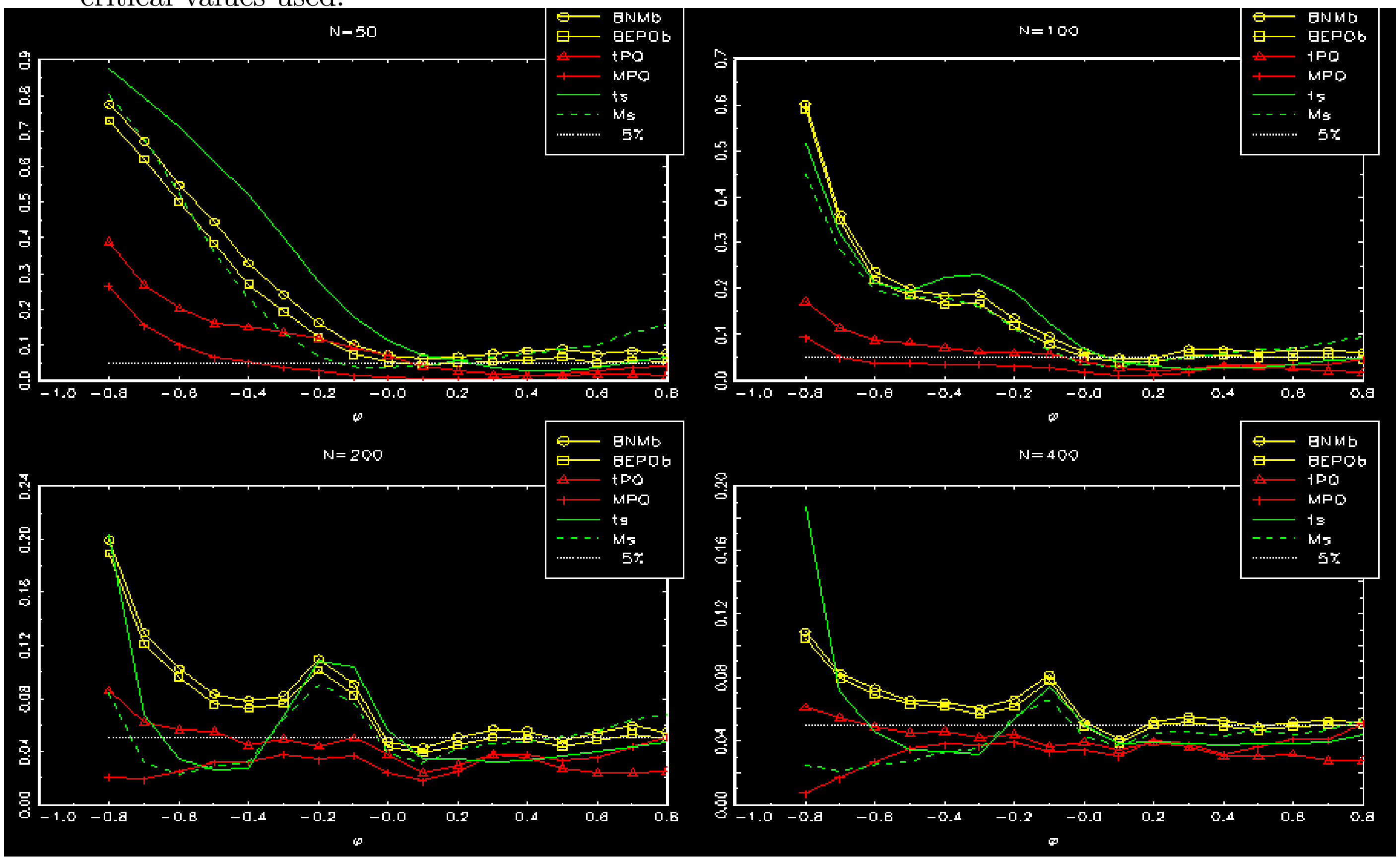


Figure 3.8. Power for model with an intercept only, $\phi = 0$. Asymptotic critical values used

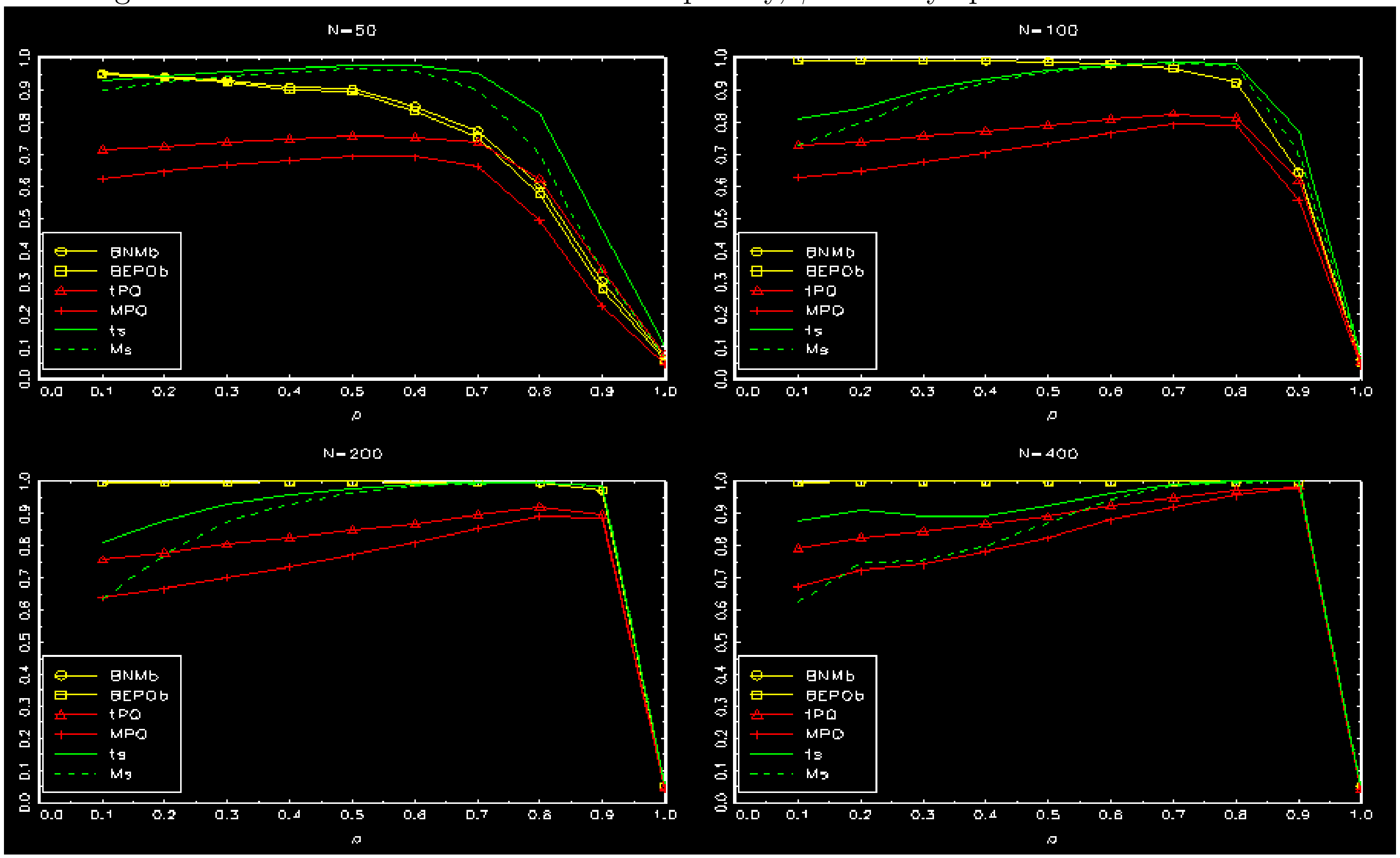
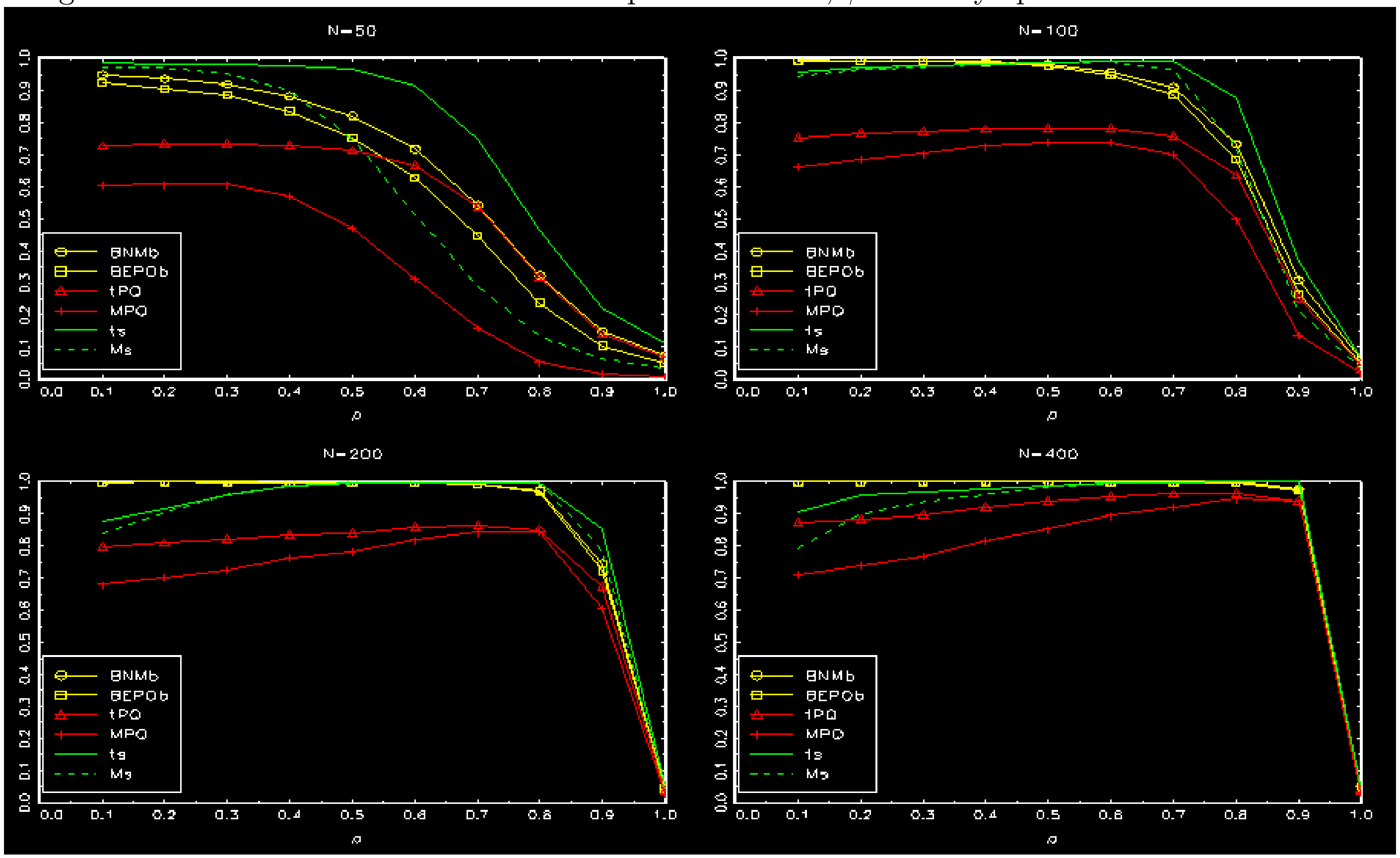


Figure 3.9. Power for model with an intercept and a trend, $\phi = 0$. Asymptotic critical values used.



CHAPTER 4

Robust Econometric Inference for Stock Return Predictability

4.1. Introduction

A fundamental issue in asset pricing is whether future stock returns are predictable using current publicly available information. The semi-strong form of efficient market hypothesis (Fama, 1970) suggests that this is not possible. Nevertheless, the seminal studies of Keim and Stambaugh (1986), Fama and French (1988) and Campbell and Shiller (1988) casted doubt on this traditional assumption and empirically demonstrated that certain financial variables have significant predictive ability over future stock returns. The existence of predictability necessarily modifies standard procedures in asset pricing, portfolio choice and performance evaluation (see Cochrane, 1999, for an excellent discussion). Fama (1991) interpreted these results as evidence of time-varying risk premia rather than evidence against market efficiency.

The early evidence on predictability motivated a significant volume of subsequent research despite which the predictability debate remains largely unsettled. For example Lettau and Ludvigson (2001, p. 842) state that “it is now widely accepted that excess returns are predictable by variables such as dividend-price ratios, earning-price ratios, dividend-earnings ratios and an assortment of other financial indicators". Other researchers (Goyal and Welch, 2008, p. 1505) remain sceptical claiming that the "profession has yet to find some variable that has meaningful and robust empirical equity premium forecasting power both in-sample and out-of-sample".

Empirical justification of arguments provided in favour of or against predictability relies on statistical inference on a set of predictive regressions and qualitative features of associated hypothesis tests such as size and power assume fundamental importance. The most common problem that undermines confidence in the reliability of predictability tests is uncertainty about the time series properties of the predictive variables. Whether standard t-tests or more sophisticated methods are employed, the quality of inference will be conditional upon correct specification of the time series properties of the predictive regressors.

A series of papers, reviewed in Campbell and Yogo (2006), have recognised that most predictability tests use financial variables that are persistent enough to be modelled as unit root or local to unity processes and are highly correlated with stock returns. Such processes assume the form of a first order autoregression with root of the form $\rho = 1 + c/n$, where n denotes the sample size. In this case standard least squares t-tests are no longer applicable for hypothesis testing (Elliott and Stock, 1994; Stambaugh, 1999) and cointegration methods need to be employed. However, the use of explanatory variables that exhibit persistence but are not necessarily pure random walks raises serious technical complications in predictive regression and cointegration methodology. Considering for example the simplest possible bivariate system

$$
\begin{aligned}
y_t &= \theta x_{t-1} + u_{0t} \\
x_t &= \rho x_{t-1} + u_{xt}
\end{aligned}
$$

with a local to unity root $\rho = 1 + c/n$ and innovations u_{0t}, u_{xt} that exhibit long run correlation the t statistic for testing the null hypothesis of no predictability, $\theta = 0$, has the following limit distribution:

$$
T_n = \frac{\left(\sum_{t=1}^{n} x_{t-1}^2\right)^{1/2}\left(\hat{\theta}_n - \theta\right)}{\hat{\Omega}_{00}^{1/2}} \Rightarrow \frac{1}{\Omega_{00}^{1/2}} \frac{\int_0^1 J_c^x(t)\, dB_0(t) + \Lambda_{0x}}{\left\{\int_0^1 J_c^x(t)^2\, dt\right\}^{1/2}} \tag{4.1}
$$

where $\Lambda_{0x} = \sum_{h=1}^{\infty} E\left(u_{0t}u_{xt-h}\right)$ and $\Omega_{00} = \sum_{h=-\infty}^{\infty} E\left(u_{0t}u_{0t-h}\right)$ denote long run covariances, $B_0(t)$ and $B_x(t)$ are Brownian motions with variances Ω_{00} and Ω_{xx} respectively and $J_c^x(t) = \int_0^t e^{c(t-s)} dB_x(s)$ the Ornstein-Uhlenbeck process associated with B_x. The distinguishing feature of local to unity limit theory for the t-statistic is that long run endogeneity cannot be removed by standard cointegration methods such as the fully modified least squares method of Phillips and Hansen (1990) or the approaches of Saikkonen (1991) and Stock and Watson (1993) that apply when the regressor is a pure random walk ($c = 0$). As pointed out by Elliott (1998), such endogeneity corrected estimators for θ would have the following asymptotic behaviour:

$$n\left(\check{\theta}_n - \theta\right) \Rightarrow \psi + c\Omega_{0x}\Omega_{xx}^{-1}$$

where ψ is a centred mixed Gaussian random variable and c is the scaling factor of the local to unity root. Since c cannot be consistently estimated, no endogeneity correction based on the above method can deliver an asymptotically mixed Gaussian estimator for θ. Analogous problems arise when the regressor has a root belonging to a larger neighbourhood of unity than local to unity roots, i.e. $\rho = 1 + c/n^{\alpha}$ where $\alpha \in (0,1)$. This class of "mildly integrated" processes was introduced by Phillips and Magdalinos (2007).

Since standard cointegration methods cannot accommodate the presence of local to unity roots in predictive regressions, a series of papers by Cavanagh, Elliott and Stock (1995), Torous, Valkanov and Yan (2004) and Campbell and Yogo (2006) have employed methods based on (4.1) where, rather than removing the additional endogeneity induced by local to unity roots, they incorporate this endogeneity in the testing procedure by constructing Bonferroni type tests. This is the current state of the art methodology for testing the predictability of stock returns.

Practical implementation of the above methodology as a tool for applied researchers presents three main drawbacks: first, the method is invalid if the regressor contains non-stationary components that are less persistent than local to unity processes, such as mildly integrated time series. Hence, each time series in the predictive regressions has to be at least as persistent as a local to unity process.

Second, incorporation of endogeneity in the inference procedure results to tests with limiting distributions that depend on a nuisance parameter that cannot be estimated, namely the scaling factor c of the local to unity process. As a result, critical values for this type of test statistics have to be chosen from a family of limit distributions by means of a Bonferroni type confidence interval on the localising coefficient c. Finally, because of the problems associated with multidimensional confidence interval construction, the above analysis is restricted to the case of a scalar regressor, i.e. a single predictive variable.

In recent work, Magdalinos and Phillips (2009) and Phillips and Magdalinos (2009), hereafter referred to as MP and PM respectively, present results that provide a framework of limit theory that can be used to validate inference in cointegrating models with regressors whose time series characteristics fall into the very general class of processes having roots in arbitrary neighbourhoods of unity. The persistence properties of these regressors may range from "near-stationarity" of mildly integrated processes to pure nonstationarity of unit root processes. Large sample endogeneity is completely removed by means of a new instrumental variables procedure, called IVX estimation. In contrast to conventional instrumental variable estimation, IVX does not use exogenous information and instruments are constructed by direct filtering of the regressor variable. The key idea behind successful endogeneity correction is explicit control of the degree of persistence of IVX instruments which are restricted within the class of mildly integrated processes.

The resulting approach yields standard chi-squared inference for testing general restrictions on a multivariate system of predictive regressions. The dimensionality of the system of predictive regressions is of particular importance for applied research since hypothesis tests on a multivariate system allow the researcher to assess the combined effects of different explanatory variables to stock returns rather than the individual effect of each explanatory variable. Given a set of explanatory variables, while each variable may have limited predictive value, there may be combinations with substantial predictive power.

The contribution of this Chapter is twofold. First, we extend the validity of the IVX methodology by allowing for the inclusion of an intercept in the predictive regression. The results of MP and PM are generalised in the above direction accommodating the modelling framework used in most applied research on stock return predictability. The second contribution consists of an empirical application of the IVX methodology in order to assess the predictive power of explanatory variables that are commonly considered as potential predictors of stock returns in applied literature. Using univariate regressions we find that the inference resulting from our methodology differs substantially from the standard least squares methodology. We decompose the market portfolio into ordered sub-categories firstly according to its size and secondly according to its book to market value. In general, we find that predictability is stronger for comparatively smaller size portfolios and for larger book to market portfolios. This shows that such decomposition of the market portfolio is meaningful and that aggregation leads to the loss of important information.

A further important contribution of our empirical analysis appears with the introduction of multivariate systems of predictive regressions. The IVX methodology allows for joint inference on combinations of different sets of both explanatory and explained variables. The importance of joint inference is highlighted in our results due to the fact that some of the regressors that appear to be insignificant in the context of univariate regressions, turn out to be jointly significant. Moreover, we test for the predictability of a set of dependent variables (decomposed portfolios with respect to size at first instance and then with respect to book to market value) by a single regressor. In other words, the methodology employed in this Chapter can be utilised for the purposes of examining whether a regressor can be a predictor of a number dependent variables simultaneously.

The final part of the empirical analysis consists of a robustness control of our empirical conclusions. First, sub-sample regressions are employed in order to assess whether empirical conclusions on the existence of stock return predictability present variations over different time periods. Second, we examine the sensitivity of our empirical conclusions to the implementation of the IVX method by conducting IVX hypothesis tests for different combinations of instrument persistence control and bandwidth selection for non parametric long run covariance estimation. The important issue of asymptotically joint optimal selection of IVX instruments and bandwidth truncation lag is addressed in Remark 4.2(c).

The proposed methodology has the potential to improve hypothesis testing with predictive regressions both by extending the range of testable hypotheses and by robustifying inference with respect to misspecification of regressor persistence. Successful implementation can shed new light on whether future bond returns, interest rates and stock returns are predictable given a publicly available information set and minimise the risk of distorted inference due to incorrect time series modelling.

The Chapter is organized as follows. Section 4.2 presents some theoretical aspects of IVX inference in systems of predictive regressions. Particular attention is devoted to accommodating the presence of an intercept in the model and deriving the relevant IVX limit theory. Section 4.3 lists the variables used in the empirical part. In Section 4.4 we apply the IVX methodology on the dataset by running individual and joint hypothesis tests. In Section 4.5 a sensitivity analysis of the inference drawn from IVX method is provided. Section 4.6 contains some concluding remarks. All proofs are included in the technical Appendix of Section 4.7. Tables and figures are presented in the last section of the Chapter.

4.2. Predictive regressions in the general vicinity of unity and IVX estimation

As is often emphasised in empirical work, economic and financial time series seem to exhibit persistence characteristics that do not always conform to the I(0)-I(1) dichotomy. In practice this means that economists wish to model persistence in regressions through series that have autoregressive roots in a general neighbourhood of unity rather than exactly at one. The primary aim of the present Chapter is to accommodate this natural relaxation of the form of nonstationarity in individual time series in the context a multivariate system of predictive regressions. In particular, we seek to extend the validity of the IVX methodology of PM to systems that include an intercept in the model and have a predictive contemporaneous structure.

We consider the following multivariate system of predictive regressions with regressors containing explanatory variables with arbitrary degree of persistence:

$$y_t = \mu + Ax_{t-1} + u_{0t}, \tag{4.2}$$

$$x_t = R_n x_{t-1} + u_{xt}, \tag{4.3}$$

where A is an $m \times K$ coefficient matrix and

$$R_n = I_K + \frac{C}{n^\alpha} \quad \text{for some } \alpha > 0, \tag{4.4}$$

and some matrix $C = \mathrm{diag}(c_1, ..., c_K)$, with $c_i \leq 0$ for all $i \in \{1, ..., K\}$. Following PM, we assume that regressor x_t in (4.3) belongs to one of the following classes of persistent processes:

P(i) *Integrated regressors, if* $C = 0$ *or* $\alpha > 1$ *in (4.4).*

P(ii) *Local to unity regressors, if* $C < 0$ *and* $\alpha = 1$ *in (4.4).*

P(iii) *Mildly integrated regressors, if* $C < 0$ *and* $\alpha \in (0, 1)$ *in (4.4).*

The aim of IVX methodology is to provide valid inference on A when there is uncertainty on the degree of persistence of the explanatory variables, i.e. there is no *a priori* knowledge of whether x_t belongs to class P(i), P(ii) or P(iii). As in PM, this is possible for all $\alpha > 1/2$.

The innovations u_{0t} and u_{xt} are assumed to be correlated linear processes. We impose an identical correlation structure to that considered in PM by considering a common Wold representation for u_{0t} and u_{xt}:

$$u_t := \begin{bmatrix} u_{0t} \\ u_{xt} \end{bmatrix} = \sum_{j=0}^{\infty} F_j \varepsilon_{t-j}, \tag{4.5}$$

where $(\varepsilon_t)_{t\in\mathbb{Z}}$ is a sequence of independent and identically distributed $(0, \Sigma)$ random vectors satisfying $\Sigma > 0$ and the moment condition $E\left\|\varepsilon_1\right\|^4 < \infty$, and $(F_j)_{j\geq 0}$ is a sequence of constant matrices satisfying $F_0 = I_{m+K}$ and

$$\sum_{j=0}^{\infty} j \left\|F_j\right\| < \infty, \tag{4.6}$$

where $\|\cdot\|$ denotes spectral norm. In accordance to standard notation, we let $F(z) = \sum_{j=0}^{\infty} F_j z^j$, and assume that $F(1) = \sum_{j=0}^{\infty} F_j$ has full rank.

As in PM, the system may be initialised at some x_0 that could be any constant or a random process $x_0(n) = o_p\left(n^{(\alpha\wedge 1)/2}\right)$ with α specified by the three cases P(i)-P(iii) listed above.

As discussed in the Introduction, endogeneity in the estimation of the coefficient matrix A in (4.2) is intimately related to the long run correlation between the innovations of the model and those of the regressor. Following standard notational convention, we denote the long run covariance matrices associated with u_t by:

$$\Omega = \sum_{h=-\infty}^{\infty} E\left(u_t u_{t-h}'\right) = F(1)\,\Sigma F(1)' \tag{4.7}$$

$$\Lambda = \sum_{h=1}^{\infty} E\left(u_t u_{t-h}'\right) \tag{4.8}$$

and $\Delta = \Lambda + E(u_1 u_1')$. In order to identify the various autocorrelation and cross correlation effects of u_{0t} and u_{xt} we consider the following partitioned forms of the matrices in (4.7) and (4.8) conformable to $u_t = (u_{0t}', u_{xt}')'$ in (4.5):

$$F(1) = \begin{bmatrix} F_0(1) \\ F_x(1) \end{bmatrix}$$

where $F_0(1)$ and $F_x(1)$ are $m \times (m+K)$ and $K \times (m+K)$ matrices respectively, and

$$\Omega = \begin{bmatrix} \Omega_{00} & \Omega_{0x} \\ \Omega_{x0} & \Omega_{xx} \end{bmatrix} \quad \text{and} \quad \Lambda = \begin{bmatrix} \Lambda_{00} & \Lambda_{0x} \\ \Lambda_{x0} & \Lambda_{xx} \end{bmatrix}. \tag{4.9}$$

In recent work, Phillips and Magdalinos (2009) have introduced a method that achieves endogeneity and bias correction in the estimation of triangular systems and is robust to the degree of regressor persistence belonging to cases (i)-(iii) above. The main idea behind the method is the construction of mildly integrated instruments by differencing the regressor x_t:

$$\Delta x_t = u_{xt} + \frac{C}{n^{\alpha}} x_{t-1}.$$

Despite the fact that the difference Δx_t is not an innovation unless the regressor is a random walk, it behaves asymptotically as an innovation after linear filtering by a matrix consisting of moderate to unity roots. Choosing an artificial matrix

$$R_{nz} = I_K + \frac{C_z}{n^{\beta}}, \quad \beta \in (0,1)\,, \quad C_z < 0, \tag{4.10}$$

IVX instruments $\tilde{z}_t$ are constructed as a first order autoregressive process with autoregressive matrix R_{nz} and innovations Δx_t:

$$\tilde{z}_t = R_{nz} \tilde{z}_{t-1} + \Delta x_t, \tag{4.11}$$

or, equivalently under zero initialisation,

$$\tilde{z}_t = \sum_{j=1}^{t} R_{nz}^{t-j} \Delta x_j. \tag{4.12}$$

The main result of PM is that, under a relatively mild assumption ($\alpha > 1/2$) that prevents the degree of regressor persistence to reach too close to stationarity, a bias-corrected two stage least squares estimator of A based on the IVX instruments $\tilde{z}_t$ is asymptotically mixed Gaussian and yields robust chi-squared inference.

In the present Chapter, we begin by extending the IVX estimation method to the case where an intercept is present in the model. This consideration is motivated by the applied literature on predictive regressions. To this end, we denote sample averages by

$$\bar{y}_n = n^{-1}\sum_{t=1}^{n} y_t, \quad \bar{u}_{0,n} = n^{-1}\sum_{t=1}^{n} u_{0t}$$

$$\bar{x}_{n-1} = n^{-1}\sum_{t=1}^{n} x_{t-1} \text{ and } \bar{z}_{n-1} = n^{-1}\sum_{t=1}^{n} \tilde{z}_{t-1}.$$

Following the notation in PM, but noting the predictive regression structure of (4.2), we construct the data, instrument and innovation matrices as follows:

$$Y = \begin{bmatrix} y_1' \\ ... \\ y_n' \end{bmatrix}, \quad U_0 = \begin{bmatrix} u_{01}' \\ ... \\ u_{0n}' \end{bmatrix} \tag{4.13}$$

$$X = \begin{bmatrix} x_0' \\ ... \\ x_{n-1}' \end{bmatrix} \text{ and } \tilde{Z} = \begin{bmatrix} \tilde{z}_0' \\ ... \\ \tilde{z}_{n-1}' \end{bmatrix}. \tag{4.14}$$

The presence of the intercept in the model can be incorporated to the IVX method by using a standard trick: Since $\bar{y}_n = \mu + A\bar{x}_{n-1} + \bar{u}_{0,n}$, (4.2) yields

$$y_t - \bar{y}_n = A\left(x_{t-1} - \bar{x}_{n-1}\right) + u_{0t} - \bar{u}_{0,n}$$

so, letting $Y_t = y_t - \bar{y}_n$, $X_t = x_t - \bar{x}_{n-1}$ and $U_{0t} = u_{0t} - \bar{u}_{0,n}$, (4.2) can be rewritten as

$$Y_t = AX_{t-1} + U_{0t}. \tag{4.15}$$

We can now proceed with IVX estimation of A from the regression equation (4.15) instead of (4.2). Accordingly, we define the new regression matrices

$$\underline{Y} = \begin{bmatrix} Y_1' \\ ... \\ Y_n' \end{bmatrix} = Y - \begin{bmatrix} \bar{y}_n' \\ ... \\ \bar{y}_n' \end{bmatrix}$$

$$\underline{X} = X - \begin{bmatrix} \bar{x}_{n-1}' \\ ... \\ \bar{x}_{n-1}' \end{bmatrix}, \quad \underline{U}_0 = U_0 - \begin{bmatrix} \bar{u}_{0,n}' \\ ... \\ \bar{u}_{0,n}' \end{bmatrix}.$$

where X, Y, U_0 and $\tilde{Z}$ are defined in (4.13) and (4.14).

Our IVX estimator then becomes:

$$\tilde{A}_n = \left(\underline{Y}'\tilde{Z} - n\hat{\Lambda}_{0x}\right)\left(\underline{X}'\tilde{Z}\right)^{-1} \tag{4.16}$$

and

$$\tilde{A}_n - A = \left(\underline{U}_0'\tilde{Z} - n\hat{\Lambda}_{0x}\right)\left(\underline{X}'\tilde{Z}\right)^{-1} \tag{4.17}$$

where $\hat{\Lambda}_{0x}$ is a non parametric estimator of $\Lambda_{0x} = \sum_{h=1}^{\infty} E\left(u_{0t}u_{x,t-h}'\right)$ based on OLS residuals from (4.3) and (4.4). The construction and properties of this non-parametric estimator is discussed below. The estimator is analogous to the FM-OLS estimator (Phillips and Hansen, 1990) in terms of its built-in bias correction term, but unlike FM-OLS there is no need for an endogeneity correction as the

IVX estimator is asymptotically mixed Gaussian because of the use of the mildly integrated instruments.

Note that the bias correction term of the IVX estimator in (4.16) involves an estimator of Λ_{0x} rather than $\Delta_{0x} = \Lambda_{0x} + E\left(u_{0t}u'_{x,t}\right)$ as in PM. This modification is due to the predictive regression structure of the model in (4.2), i.e. the fact that y_t is regressed upon x_{t-1} rather than x_t. Note also that the estimator does not involve a demeaned version of the matrix of instruments as the IVX estimator in (4.16) is exactly invariant to demeaning $\tilde{Z}$ by $\bar{z}_{n-1}$.

We now discuss the issue of non-parametric estimation of Λ_{0x}. Letting

$$\underline{A} = [\mu, A] \quad \text{and} \quad \underline{X}_t = (1, x'_t)' \tag{4.18}$$

we can write (4.2) as

$$y_t = \underline{A}\ \underline{X}_{t-1} + u_{0t}. \tag{4.19}$$

This yields the following OLS estimator of $\underline{A}$:

$$\tilde{\underline{A}}_{OLS} = \left(\sum_{t=1}^{n} y_t \underline{X}'_{t-1}\right)\left(\sum_{t=1}^{n} \underline{X}_{t-1}\underline{X}'_{t-1}\right)^{-1}. \tag{4.20}$$

The presence of the the intercept in (4.2) is taken into account by constructing the estimated residuals of the model from (4.19):

$$\tilde{u}_{0t} = y_t - \tilde{\underline{A}}_{OLS}\underline{X}_{t-1} \tag{4.21}$$

where $\underline{X}_t$ and $\underline{\tilde{A}}_{OLS}$ are given by (4.18) and (4.20) respectively. Since there is no intercept in (4.3), the estimated regressor residuals can be obtained in a standard way:

$$\hat{u}_{xt-h} = x_{t-h} - \hat{R}_n x_{t-h-1} \tag{4.22}$$

where $\hat{R}_n$ is the OLS estimator of R_n in (4.3). Given the estimated residuals in (4.21) and (4.22) we can consistently estimate the long run covariance Λ_{0x} by the following Newey-West type HAC estimator:

$$\hat{\Lambda}_{0x} = \frac{1}{n} \sum_{h=1}^{M} \left(1 - \frac{h}{M+1}\right) \sum_{t=h+1}^{n} \tilde{u}_{0t} \hat{u}'_{xt-h} \tag{4.23}$$

where M is a bandwidth parameter satisfying $M \to \infty$ as $n \to \infty$. The estimator of Λ_{00} is constructed in a similar manner, by replacing $\hat{u}_{xt-h}$ in (4.23) by $\hat{u}_{0t-h}$. The above estimators have standard consistency properties. Consistency of $\hat{\Lambda}_{00}$ (and hence of $\hat{\Omega}_{00}$) is enough for the requirements of IVX limit theory. On the other hand, $\hat{\Lambda}_{0x}$ is part of a first order bias correction on the IVX estimator so, in view of Theorems 3.4 and 3.7 of PM (see also Theorem 4.1 below), its consistency rate should ensure that the condition

$$n^{\frac{1-(\alpha \wedge \beta)}{2}} \left(\hat{\Lambda}_{0x} - \Lambda_{0x}\right) \to_p 0 \tag{4.24}$$

is satisfied. The exact consistency rate of $\hat{\Lambda}_{0x}$ is given in the following result.

Lemma 4.1.

(i) *Let u_t be a linear process given by (4.5) satisfying (4.6) and* $E\left\|\varepsilon_1\right\|^4 < \infty$. *Then, for all* $\alpha > 1/2$,

$$\hat{\Lambda}_{0x} - \Lambda_{0x} = O_p\left(\max\left\{\frac{M}{n^{1/2}}, \frac{1}{M}\right\}\right).$$

(ii) *Let* $M = L(n)\, n^{\gamma}$ *for some slowly varying function* L *and* $\gamma > 0$. *Choose* $\beta \in (2/3, 1)$. *Then a choice of* $\gamma = 1/4$ *guarantees the validity of (4.24) for any* $\alpha > 1/2$. *If* $\alpha > 2/3$, *(4.24) holds under the optimal choice of bandwidth* $\gamma = 1/3$.

The proof of Lemma 4.1 is given in Section 7.

Obtaining a limit theory for the modified IVX estimator in (4.16) can be achieved by using similar methods as in the PM paper. It turns out that effect of the presence of an intercept in the model is manifest only on the limit distribution of the $\underline{X}'\tilde{Z}$. As a result, asymptotic mixed normality of the IVX estimator continues to apply and the intercept affects only the form of the limiting random variance. The main result is presented in the following theorem and is comparable with Theorems 3.4 and 3.7 of PM. All steps of the proof associated with the presence of the intercept (and hence not covered by PM) are included in a sequence of lemmata in Section 4.7.

Theorem 4.1. *Consider the model (4.2) - (4.4) with instruments* $\tilde{z}_t$ *defined by (4.12). Then, the following limit theory applies for the estimator* $\tilde{A}_n$ *in (4.16):*

(i) *If* $1/2 < \beta < \min(\alpha, 1)$*:*

$$n^{\frac{1+\beta}{2}} \operatorname{vec}\left(\tilde{A}_n - A\right) \Rightarrow MN\left(0, \left(\tilde{\Psi}_{xx}^{-1}\right)' C_z V_{zz} C_z \tilde{\Psi}_{xx}^{-1} \otimes \Omega_{00}\right),$$

as $n \to \infty$*, where*

$$\tilde{\Psi}_{xx} = \begin{cases} \Omega_{xx} + \int_0^1 \underline{B}_x dB_x' & \textit{under P(i)} \\ \Omega_{xx} + \int_0^1 \underline{J}_C dB_x' + \int_0^1 \underline{J}_C \underline{J}_C' ds C & \textit{under P(ii)} \\ \Omega_{xx} + V_{xx} C & \textit{under P(iii)} \end{cases},$$

B_x *is a Brownian motion with variance* Ω_{xx} *and* J_C *the associated Ornstein-Uhlenbeck process,*

$$\underline{B}_x(t) = B_x(t) - \int_0^1 B_x(t)\, dt,\ \ \underline{J}_C(t) = J_C(t) - \int_0^1 J_C(t)\, dt,$$

$$V_{xx} = \int_0^\infty e^{rC} \Omega_{xx} e^{rC} dr \ \text{ and } \ V_{zz} = \int_0^\infty e^{rC_z} \Omega_{xx} e^{rC_z} dr.$$

(ii) *If* $\alpha \in (1/2, \beta)$ *then Theorem 3.7 of PM continues to apply.*

Remarks 4.1.

(a) Theorem 4.1 shows that the presence of an intercept in the model does not affect the main asymptotic property of IVX estimation, mixed Gaussianity. Asymptotic bias and endogeneity removal are achieved under the same restriction ($\alpha > 1/2$) as the original PM paper.

(b) A comparison between Theorems 3.4 and 3.7 of PM and Theorem 4.1 above shows that the effect of the intercept on the limiting distribution of the IVX estimator depends on the degree of regressor persistence. For local to unity and unit root processes, this effect is manifest on the limit distribution of the $n^{-(1+\beta)}X'\tilde{Z}$ matrix, where the Brownian motion B_x and the Ornstein-Uhlenbeck process J_C are replaced by their demeaned counterparts $\underline{B}_x$ and $\underline{J}_C$ respectively. IVX limit theory remains unaffected by the presence of an intercept in the case of mildly integrated regressors.

Let $\bar{z}_{n-1} = n^{-1}\sum_{t=1}^{n}\tilde{z}_{t-1}$,

$$\underline{\tilde{Z}} = \tilde{Z} - \begin{bmatrix} \bar{z}'_{n-1} \\ \ldots \\ \bar{z}'_{n-1} \end{bmatrix} \quad \text{and} \quad P_{\underline{\tilde{Z}}} = \underline{\tilde{Z}}\left(\underline{\tilde{Z}}'\underline{\tilde{Z}}\right)^{-1}\underline{\tilde{Z}}' \tag{4.25}$$

denote theprojection matrix to the column space of the demeaned instrument matrix $\underline{\tilde{Z}}$. The mixed normal limit theory of Theorem 4.1 implies that linear restrictions on the cointegrating coefficients A generated by (4.2) can be tested by a standard Wald test. In particular,

$$H_0 : H\text{vec}\,(A) = h, \tag{4.26}$$

where H is a known $r \times mK$ matrix with rank r and h is a known vector, may be tested using the Wald statistic

$$W_n = \left(H\text{vec}\tilde{A}_n - h\right)'\left[H\left\{\left(\underline{X}'P_{\underline{\tilde{Z}}}\underline{X}\right)^{-1} \otimes \hat{\Omega}_{00}\right\}H'\right]^{-1}\left(H\text{vec}\tilde{A}_n - h\right) \tag{4.27}$$

where $P_{\underline{\tilde{Z}}}$ is defined in (4.25), $\tilde{A}_n$ is the IVX estimator in (4.16) and $\hat{\Omega}_{00}$ is a consistent non parametric estimator of Ω_{00} in (4.9).

Theorem 4.2. *Under the null hypothesis (4.26) of general linear restrictions on A, the Wald statistic in (4.27) has the following limit distribution: $W_n \Rightarrow \chi^2(r)$ for every $\alpha > 1/2$.*

Remarks 4.2.

(a) Theorem 4.2 is an immediate corollary of the mixed Gaussian limit theory for the IVX estimator of Theorem 4.1. It shows that the IVX-based Wald test in (4.27) can provide an inference procedure that is robust to a wide range of persistent data generating processes, ranging from mildly integrated processes to pure random walks. It is hoped that this procedure will provide a unifying framework for hypothesis testing in predictive regressions which maintains good statistical properties under misspecification of the time series characteristics of the regressors.

(b) As in the original PM paper, the validity of Theorem 4.2 is restricted by excluding regressors that contain roots close to the boundary with stationarity. This limitation of the IVX method is intimately related to its feasibility: since the IVX instruments in (4.12) are constructed from the regressors without imposing any exogenous orthogonality assumption, moving towards the stationary region increases the effect of simultaneity bias and, eventually, makes estimation impossible. It is well known that when x_t is a stationary process ($\alpha = 0$), the system (4.2)-(4.4) cannot be identified (or of course estimated) without exogenous information in the form of instruments that satisfy an orthogonality and a relevance condition. For mildly integrated systems, a calculation of simultaneity bias appears in MP for all $\alpha \in (0,1)$. This bias takes a simpler form for $\alpha > 1/3$ in which case $n^{-\frac{1+\alpha}{2}} \sum_{t=1}^{n} \left(u_{0t} x'_{t-1} - \Lambda_{0x} \right)$ has a centred normal limit distribution. The more stringent restriction $\alpha > 1/2$ is needed for for controlling the estimation error in the non parametric estimation bias correction in the above sum, i.e. ensuring that $\hat{\Lambda}_{0x} - \Lambda_{0x}$ satisfies (4.24).

(c) Implementation of the method requires a choice for β for the construction of the IVX instruments. As explained in PM, it is recommended to choose β from the interval $(2/3, 1)$. Such a choice allows mean squared error (MSE) efficient non parametric estimation of the long run covariance matrix Λ_{0x} for unit root and local to unity regressors as well as mildly integrated regressors with $\alpha > 2/3$. Recent work in progress by PM suggests that an asymptotic MSE minimising choice of β is given by

$$\beta = \frac{1 + 2\gamma}{2}, \tag{4.28}$$

where γ is the polynomial rate of growth of the bandwidth parameter $M = L(n) n^{\gamma}$ of the Newey West estimator of the long run covariance matrix Λ_{0x} in (4.23). Hence there is an one to one correspondence between optimal asymptotic MSE choice of IVX instruments and optimal asymptotic MSE non-parametric estimation of Λ_{0x}. Since the optimal rate of bandwidth growth for the Bartlett kernel employed in (4.23) is $n^{1/3}$, substituting $\gamma = 1/3$ in (4.28) yields a choice $\beta = 5/6$. We employ this choice of β in the subsequent empirical analysis.

(d) Note that demeaning the instrument matrix $\tilde{Z}$ in the Wald statistic produces finite sample gain: The conclusion of Theorem 4.1 (i) without demeaning $\tilde{Z}$ can be written informally as:

$$n^{\frac{1+\beta}{2}} \operatorname{vec}\left(\tilde{A}_n - A\right) \Rightarrow MN\left(0, \lim_{n\to\infty}\left(\frac{1}{n^{1+\beta}}\underline{X}' P_{\tilde{Z}}\underline{X}\right)^{-1} \otimes \Omega_{00}\right). \tag{4.29}$$

The identity (4.56) in Section ?? implies that $\left(\underline{X}' P_{\underline{\tilde{Z}}}\underline{X}\right)^{-1} \leq \left(\underline{X}' P_{\tilde{Z}}\underline{X}\right)^{-1}$ in the positive semidefinite sense, so there is finite sample efficiency gain associated with demeaning the instrument matrix. This gain dissapears asymptotically and the Wald statistic with and without demeaning has the same chi-squared distribution, as (4.53) in Section ?? shows.

4.3. The Dataset

We employ two datasets for the predictability tests we conduct in the following section. The sample period for both datasets is January 1927 to December 2007. The first dataset contains the stock portfolio returns used as dependent variables. The source for these portfolios' returns is the widely used Kenneth French's online data library[1]. In particular, U.S. market returns are proxied by the Center for Research in Security Press (CRSP) value weighted returns. Moreover, we employ monthly value-weighted returns of ten portfolios formed on the basis of stocks' market value (Size portfolios) and monthly value-weighted returns of ten portfolios sorted according to stocks' book equity to market value of equity ratio (Value portfolios). We calculate returns in excess of the corresponding 1-month T-bill rate.

The second dataset contains the variables that are commonly used as regressors in predictability tests and for which there is uncertainty for their order of integration[2]. This is an updated version of the dataset used in Goyal and Welch (2008)[3]. These 11 variables refer to:

T-bill rate (tbl): This is the 3-month US Treasury bill rate taken from the economic research database at the Federal Reserve at St. Louis (FRED). For the period before 1934 it is extracted from the NBER Macrohistory database. The T-bill rate has been used as a predictor of future stock returns *inter alia* by Pesaran and Timmermann (1995), Pontiff and Schall (1998), Torous *et al.* (2004), Campbell and Yogo (2006), Ang and Bekaert (2007), Avramov (2002) and Campbell and Thompson (2008).

[1]This library is available at
http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

[2]Our focus is on these variables because our econometric methodology is developed to deal with the inference problems arising from the uncertainty with respect to their order of integration. However, it is worth mentioning that various other variables have been used in predictive regressions of future stock returns (see Goyal and Welch, 2008 for an exhaustive list).

[3]We would like to thank Amit Goyal for providing us with this dataset.

Long-term yield (lty): This is the long-term US government bond yield from Ibbotson's Stocks, Bonds, Bills and Inflation Yearbook. The difference between the long-term yield and the T-bill rate is the **term spread (tms)**. These two variables have been widely used in predictability tests (see for example, Keim and Stambaugh (1986), Fama and French (1989), Pontiff and Schall (1998), Torous *et al.* (2004) and Campbell and Yogo (2006)).

Default yield spread (dfy): This is the difference between the BAA and AAA-rated corporate bond yields taken from FRED. Fama and French (1989), Torous *et al.* (2004), Avramov (2002) and Campbell and Thompson (2008) provide examples of studies that have employed dfy as a predictive regressor.

Dividend price ratio (d/p): This is the difference between the log of dividends and the log of stock prices. Dividends are calculated using a 12-month rolling sum of dividends paid on the S&P 500 index. The difference between the log of dividends and the log of lagged prices is the **dividend yield (d/y)**. These two variables have been the most commonly used predictors of future stock returns. An indicative list of previous studies contains Rozeff (1984), Campbell (1987), Campbell and Shiller (1988), Fama and French (1988), Hodrick (1992), Lamont (1998), Stambaugh (1999), Goyal and Welch (2003), Lewellen (2004), Torous *et al.* (2004), Lettau and Ludvigson (2005), Campbell and Yogo (2006) and Campbell and Thompson (2008).

Earnings price ratio (e/p): This is the difference between the log of earnings and the log of prices. Earnings are calculated using a 12-month rolling sum of earnings of companies listed on the S&P 500 index. Campbell and Shiller (1988), Fama and French (1988), Pesaran and Timmermann (1995), Lamont (1998), Lewellen (2004) and Campbell and Thompson (2008) are examples of studies which employed this variable.

Dividend payout ratio (d/e): This is the difference between the log of dividends and the log of earnings, as previously defined. The study of Lamont (1998) makes a convincing case for using this payout ratio as a potential predictor of future stock returns.

Book-to-Market ratio (b/m): This is the ratio of book value to market value for the Dow Jones Industrial Average. The book value is taken from Value Line's website, specifically their Long-Term Perspective Chart of the Dow Jones Industrial Average. This ratio has been used in the studies of Kothari and Shanken (1997), Pontiff and Schall (1998), Lewellen (2004), Avramov (2002) and Campbell and Thompson (2008) *inter alia*.

Net equity expansion (ntis): This is a measure of corporate issuing activity and it is calculated as the ratio of the 12-month moving sum of net equity issues by NYSE listed stocks divided by the total end-of-year market capitalization of these stocks. Net equity issuing activity refers to Initial Public Offerings, Seasoned Equity Offerings, stock repurchases minus distributed dividends and it is calculated from CRSP data (see Goyal and Welch, 2008, for details). This variable is closely related to the net payout yield that has been proposed by Boudoukh, Michaely, Richardson and Roberts (2007) as a predictor of future stock returns.

Cross-sectional premium (csp): The cross-section premium measures the relative valuations of high- and low- CAPM beta stocks and it has been employed in predictability tests by Polk, Thompson and Vuolteenaho (2006). For this particular variable, the available data start from May 1937 to December 2002.

4.4. Empirical analysis

We initiate the discussion in this Section by examining the time series properties of the data. More specifically, we run unit root tests on the series used as regressors. Three tests are employed: the Augmented Dickey Fuller (ADF) test, the DF-GLS test derived by Elliot *et al.* (1996) and the Phillips Perron (PP) test. For both ADF and DF-GLS the Bayesian information criterion is used for the determination of lag length. The results of the test statistics are presented in Table 4.1. The null hypothesis of non-stationarity is not rejected for lty, d/y, d/p across the different tests used. Strong evidence of stationarity appear for the series of the term spread. For the rest of the data series, the inference regarding the null hypothesis of non-stationarity does not remain unchanged when different test statistics are considered. This is the case, for example, for the T-bill rate which is suggested to be stationary by the DF-GLS test, and not stationary by the ADF and PP test. For the series of net equity expansion, the null of non-stationarity is strongly (at the 1% level) rejected by ADF and PP tests, but the same null hypothesis is not rejected by DF-GLS test.

The evidence provided by Table 4.1 confirm the findings of other studies concerning the uncertainty about the time series properties of the data used as predictors of stock returns, and motivate the use of local to unity framework for the examination of stock return predictability.

4.4.1. Univariate regressions

The first set of predictability tests we report refers to the case where the dependent variable is the CRSP excess return and the predictor is the lagged value of each of the 11 variables described in the previous section. In-sample predictability of aggregate market returns is the focus of the vast existing literature. Our contribution is to examine how inference is modified once we employ the proposed econometric methodology. Table 4.2 contains the results both for a standard least squares approach and the new IVX estimation approach. Moreover, it reports the long-run correlation coefficient of the residuals from regression models (4.2) and (4.3).

The least squares approach would point to the conclusion that the null hypothesis of no predictability can be rejected (at a level lower than 5%) when the lagged series of the earning-price ratio, the cross-sectional premium and the net equity expansion are alternatively employed as predictors. The ability of T-bill rate to predict future stock returns is on the borderline of statistical significance. There is some very weak evidence in favour of the predictive ability of the dividend yield, dividend-price ratio, book to market value ratio and term spread, while there is no such evidence for the dividend payout ratio, long-term yield and default yield to be regarded as predictors of next month excess market returns. The signs of the coefficients are in accordance with the existing literature. An increase in the d/y, d/p, e/p, b/m and tms would be associated with an increase in next period's excess market returns, while an increase in the net equity expansion and the T-bill rate would be associated with negative future returns.

These standard results demonstrate that the overall evidence on short-term predictability is very weak, hence the debate is still wide open. Moreover, a least squares approach is well known to yield biased estimates in the presence of highly persistent regressors (Stambaugh, 1999), especially when the residuals of the system's regressions are highly correlated. We indeed find that the residuals' long-run correlation is particularly high when the dividend yield, dividend-price ratio, book to market value ratio, earnings-price ratio and default yield are used as regressors. This pattern invalidates the inference based on standard least squares and motivates the examination of the results from the IVX approach.

For each of the employed regressors we report in Table 4.2 the estimated IVX coefficient as well as the corresponding Wald statistic to test the null hypothesis that this coefficient is zero. It should be mentioned that the estimated coefficients are not directly comparable with the ones derived from least squares because they are derived by using the instrumental variable $\tilde{z}_{t-1}$, constructed in equation (4.12).

Moreover, the reported results refer to the case where $\beta = 5/6$ and $\gamma = 1/3$. In the following section we perform a rigorous sensitivity analysis to examine how the choice of these values affects inference.

Using the IVX approach, we derive striking conclusions. Five variables (dividend yield, dividend-price ratio, earnings-price ratio, book to market value ratio and net equity expansion) are now found to predict excess market returns at the 5% level or lower. In particular, the strongest evidence in favour of predictability is documented for net equity expansion and the book to market value ratio. On the other hand, the previous evidence on the significance of the T-bill rate and the cross sectional premium as a predictors is overturned when our methodology is employed. With respect to the signs of the estimated coefficients on the constructed $\tilde{z}_{t-1}$ variable, the standard arguments carry through. A positive relationship between next period excess market returns and this variable is reported for the case of the d/y, d/p, e/p and b/m, while a negative one is reported for the case of ntis. Taken as a whole, inference based on the IVX approach is considerably different from the standard least squares one, demonstrating the important role that the regressors' time series properties play. We argue that there is significant evidence supporting the case of predictability through a set of commonly used regressors, even when we take into account the uncertainty surrounding their order of integration, confirming that the market premium is time-varying indeed.

Apart from examining the issue of predictability on the aggregate market portfolio, an interesting question is whether these variables have predictive ability over components of this portfolio. This issue is worth examining for a series of reasons. Most importantly, if components of the market portfolio are predictable, this would be valuable information for the formation of optimal portfolios (Kandel and Stambaugh, 1996), performance evaluation of investment managers (Christopherson *et al.*, 1998), asset pricing models (Ferson and Harvey, 1999), and hence for the implied cost of capital for the companies in that market segment. Keim and Stambaugh (1986) tested for predictability using size-sorted portfolios, while Ferson and Harvey (1999) and Avramov (2002) employed double-sorted portfolios on the basis of size and book to market value. Most recently, Kong *et al.* (2009) provided an exhaustive analysis for size, book to market value and industry portfolios' returns using a similar set of regressors as well as lagged industry returns. However, they rely on a least squares approach and a bootstrap procedure for their inference analysis that may suffer from low power (see Campbell and Yogo, 2006). We sidestep this problem by using the proposed IVX approach to test for predictability in the returns of each of the ten size and the ten book to market value sorted portfolios, described in the previous section.

Table 4.3 contains the estimation results corresponding to the ten size portfolios while Table 4.4 contains the results for the ten book to market value results. With respect to the size portfolios, we find significant evidence in favour of predictability across all ten portfolios. However, there are considerable differences between them with regard to which regressors contain predictive ability and how strong this evidence is. The most interesting results refer to the smallest size decile. For this portfolio, we find that 7 out of the 11 considered variables exhibit significant in-sample predictive ability at the 5% level or lower. These are the d/y, d/p, b/m, ntis as well as dfy, tbl and tms. In other words, despite the fact that the last three variables were not found to be significant predictors of the aggregate market portfolio's returns, they have predictive ability over small cap stocks' returns, in line with Ferson and Harvey (1999) and Kong *et al.* (2005). The implication of this result is that small stocks' premia are time-varying indeed and that the default yield, along with the term spread and the 3-month T-bill rate can capture at least part of this time-variation. On the other hand, the S&P 500 e/p ratio is not found to be a significant predictor for the future returns of this decile portfolio, while it was previously found to be significant for the market portfolio. This finding demonstrates the importance of decomposing the market portfolio into its components for predictability tests too. The most significant predictors for small cap stocks' returns are found to be the book to market value ratio,the default yield spread and the net equity expansion (rejection of the null of no predictability at the 1% level).

As we move away from the small size decile towards the big size decile (Size 10), inference with respect to which regressors' coefficients are significantly different from zero is modified. The polar case of the biggest size decile yields some quite intriguing results. In particular, d/p, d/y, e/p, b/m and ntis are found to be significant predictors at the 5% level or lower. An inspection of the results in Table 4.3 allows us to derive the following conclusions. The b/m, d/y, d/p and ntis have significant in-sample predictive ability for all of the size-sorted portfolios' returns, as the null hypothesis of no predictability is rejected, at a 5% level or lower, for any regression combining the each of the decomposed portfolios with these four variables. The T-bill rate is a significant predictor, at the 5% level, only for the smallest size portfolio's returns, while csp, d/e and lty are not found to predict future returns for any portfolio. Apart from a few big size stock portfolios, the rest portfolios' returns are predictable through the default yield and the term spread. Finally, the earnings-price ratio has predictive ability for most but a few small size portfolios' returns. These findings demonstrate the different characteristics in the cross-section of stock returns, providing support for the argument that the size premium (i.e. the spread between small and big cap stock returns) is time-varying indeed and this time-variation exhibits a partly predictable pattern through a set of commonly used regressors that are thought to contain information for the underlying economic conditions (see the seminal study of Fama and French, 1992).

Inspecting the results in Table 4.4, for the univariate predictability tests with respect to the returns of portfolios formed on the basis of stocks' book to market value ratio, interesting cross-sectional differences emerge again. As a general rule, there is much stronger evidence in favour of predictability for value portfolios (deciles 8, 9 and 10) than for growth portfolios (deciles 1, 2 and 3). This is true both in terms of the number of regressors that are found to have predictive ability and the level of significance at which the null of no predictability is rejected. With respect to the common source of predictability, all portfolios' returns are found to be predictable through ntis and the b/m ratio at the 5% level or lower[4]. The S&P 500 earning-price ratio is also found to have strong predictive ability for most but the smallest value portfolio. As we move towards value portfolios, d/y and d/p contain statistically significant predictive ability; for growth portfolios, the null of no predictability cannot be rejected when using these two variables.

Interestingly, the default yield spread is a reliable predictor only for value portfolios' returns, supporting the argument of Fama and French (1993) that the value premium may represent compensation for distress risk, since the default yield spread tends to widen in periods of adverse economic conditions. Moreover, the term spread is a statistically significant predictor only for deep value (decile 10) stocks. On the other hand, the payout ratio (d/e), the long-term yield (lty), the T-bill rate and the cross-sectional premium (csp) are not found to predict returns for any value-sorted portfolio considered. Overall, these findings not only confirm the ability of commonly used regressors to predict future returns for value-sorted portfolios but they also point to their ability to explain cross-sectional differences in their returns. As a result, the approach of Petkova and Zhang (2005) to examine the value premium within a conditional asset pricing framework is legitimate, motivating also the use of different sets of variables.

Summarizing the previous evidence, our univariate tests show that net equity expansion (ntis) is the most reliable predictor of future stock returns across all

[4]It should be reminded that the book-to-market value ratio employed as a regressor corresponds to the Dow Jones Industrial Average stocks; it is not the b/m ratio of the stocks in each portfolio considered.

of the examined portfolios. This finding supports the arguments of Boudoukh *et al.* (2007), who introduce a closely related variable, the net payout yield. On the other hand, the dividend payout ratio suggested by Lamont (1998) and the cross-sectional premium suggested by Polk *et al.* (2006) do not contain any statistically significant information for future stock returns using our testing methodology. Regarding the rest of the variables, the dividend yield and the dividend-price ratio are found to predict next month's returns for most of the examined portfolios, with the exception of growth portfolios. Nevertheless, we should iterate the standard finding in predictability studies, that the degree of explanatory power for all of these regressors is very low. This is an expected feature given that we attempt to explain a very noisy variable, stock returns, through highly persistent regressors.

4.4.2. Multivariate regressions

The predictability literature has not focussed only on the ability of a single economic or financial variable to explain future stock returns. Actually, from the early contributions to this literature onwards (see e.g. Fama and French, 1989), multivariate regressions were employed and the joint significance of these regressors was examined, apart from their individual contribution to the explanatory ability of the model. This approach was informative for tests of the semi-strong form of market efficiency, because in its original version this form was assuming lack of predictability from any set of underlying variables, not just each variable in isolation (Fama, 1970). This approach is still very important for asset pricing tests and conditional performance evaluation of investment strategies, because in these applications the focus is on the explanatory ability and fitness of the whole regression model, not each variable on its own. Moreover, this evolving literature has documented the individual predictive ability for a series of variables, and hence for practical reasons we would like to compare the overall predictive ability of parsimonious regression models that use only subsets of the suggested variables.

Given the importance of multivariate predictability regressions, it is frustrating that the recent methodological contributions that correct for the bias in the least squares estimation are developed for single variable regressions. This is true both for studies that use asymptotic t-statistics for near-unit root regressors (Torous *et al.*, 2004 and Campbell and Yogo, 2006) and for studies that rely on conditional t-tests (Lewellen, 2004)[5]. As mentioned in Section 4.2, an appealing feature of the proposed econometric methodology is that we can overcome this limitation and devise predictability tests that use multiple regressors as well as multiple regressands. The latter case proves particularly useful for testing whether a variable or a set of variables have predictive ability over the entire cross-section of portfolio returns, e.g. the ten size or the ten value portfolios that we previously examined. We perform such tests and comment on the derived results.

[5]The Bayesian approach of Stambaugh (1999) allows for multiple regressors, but prior belief for the order of their integration is required.

Table 4.5 contains the results for multivariate regressions when the excess market return is used as the dependent variable and a subset of the previously examined highly persistent regressors are used as independent variables. It should be mentioned that we exclude from this exercise the cross-sectional premium due to the lack of data for the whole sample period. Moreover, the dividend yield has a correlation coefficient of 0.99 with the dividend-price ratio due to their construction. Therefore, these variables cannot be included in the same regression model, given the obvious multicollinearity problem; for the results we present, the dividend yield has been included[6]. In addition, only two of the T-bill rate, long-term yield and term spread can be used in the same regression, because each of them is by definition a linear combination of the other two variables and this feature would lead to a singular regressors' matrix. We examine various combinations. Firstly, we consider whether the null of no predictability can be rejected when all of these eight regressors are included in the model. We find very strong evidence (rejection of the null hypothesis of joint insignificance even at the 1%) in favour of joint predictability. This result confirms the conjecture that even when the uncertainty regarding the order of their integration is taken into account, these commonly used regressors can predict, in-sample, excess market returns over the following month. Given the previous discussion, this results points towards time-varying risk premia and confirms the ability of these variables to capture, at least partly, the evolution of these premia.

[6]Results are very similar when the dividend-price ratio is used instead.

The second step we take is to separate the variables that were found to be individually significant predictors from those that were not and run tests using various combinations within each of these two subsets. For the significant predictors this exercise can provide an understanding of which combinations yield the strongest evidence in favour of predictability, while for the individually insignificant variables, as it is standard in the case of multiple regressions, we are interested to examine whether some of their combinations can still be jointly significant in predicting excess market returns. All of the 2-, 3- and 4-variable combinations of the individually significant variables are found to be jointly significant too, strengthening the case in favour of predictability. Very strong evidence for the joint significance of the variables is reported when net equity expansion and the earnings-price ratio and/or the dividend yield are included in the model. On the other hand, for no combination of the individually insignificant variables can we reject the null of no predictability at the 5% level. Combining the default yield spread with the T-bill rate and the dividend payout ratio there is evidence for joint significance only at the 10% level. In summary, using our testing methodology there is no strong evidence supporting the predictive ability of the T-bill rate, default yield spread, dividend payout ratio and the term spread, neither individually nor jointly.

Tables 4.6 and 4.7 present the results derived from testing whether a set of portfolios' returns can be predicted by each of the 11 lagged variables we examine. In other words, this is a Wald test for the null hypothesis that the coefficients derived by regressing each portfolio's returns on the transformation $\tilde{z}_{t-1}$ of the highly persistent variable x_{t-1} are jointly equal to zero. With respect to the size portfolios, we have strong evidence against the null hypothesis of no joint predictability when the dividend payout ratio, the book to market ratio, the default yield spread and net equity expansion are employed as regressors. This evidence clearly leads to the conclusion that these variables contain significant information explaining the time-varying premia across size-ordered portfolios. For the dividend yield and the dividend-price ratio, the Wald statistic marginally fails to reject the null at the 5% level. The Wald statistic only rejects the joint null hypothesis for these regressors at the 10% level. An inspection of the estimated IVX coefficients shows a clear gradient as we move from small size portfolios (decile 1) towards big size portfolios (decile 10), with small size portfolios being more sensitive to d/y, d/p and ntis than big size portfolios. On the other hand, no evidence of predictability is reported for the long-term yield, the T-bill rate, the earnings-price ratio, the cross-sectional premium and the term spread.

With respect to the ten book to market value ordered portfolios, no evidence in favour of joint predictability is found when the dividend payout ratio, long-term yield, dividend yield, dividend-price ratio, T-bill rate, earnings-price ratio, cross-sectional premium and the term spread are individually employed as explanatory variables. More interestingly, using the Dow Jones Industrial Average book to market value ratio as a regressor we find strong evidence for predictability and there is a clear pattern with respect to the portfolios' sensitivities; value portfolios' returns are much more sensitive to this variable in comparison to growth portfolios' returns. The most reliable joint predictor appears to be the default yield spread, generating also a very clear gradient; the estimated IVX coefficient corresponding to the value portfolio (decile 10) is almost ten times greater relative to the growth portfolio (decile 1). Since the default yield spread is related to the prevailing credit conditions by capturing the evolution of default risk for corporate bonds, it is legitimate to argue that the time-varying premium across the value portfolios can be partly attributed to the default risk premium. The net equity expansion is found to be a significant predictor for all value ordered portfolios jointly. The Wald test statistic corresponding to the null of no predictability is rejected even at 1% level.

4.5. Further results

4.5.1. Sub-period analysis

The results we reported in the previous section refer to the whole sample period, from January 1927 to December 2007. It is common practice in the literature to test for predictability in sub-periods too, examining whether the whole period results carry through (see Torous *et al.*, 2004 and Campbell and Yogo, 2006 for recent examples). There are two main reasons why this exercise is informative. Firstly, the evidence in favour of predictability may simply be attributed solely to early periods when this pattern was not widely documented. This explanation essentially implies that these predictable relationships were due to market inefficiencies that later disappeared, once investors became aware of them and devised asset allocation strategies aiming at exploiting them. On the other hand, if these predictable patterns persist through time, the implication is that they reflect time-varying risk premia rather than mispricings (Fama, 1991). Secondly, there is the possibility that the degree of predictability as well as returns' sensitivities to these variables have changed through time, due to the fundamental developments in the US economy and the structure of financial markets during these 80 years. Along these lines, Viceira (1997) explicitly tests for structural breaks in the predictability relationships, while Lettau and van Nieuwerburgh (2008) allow for a time-varying relationship between expected returns and the commonly used financial ratios. Gonzalo and Pitarakis (2009) discuss the instability of the predictability hypothesis and suggest that predictability appears for some valuation ratios during periods of recession.

Table 4.8 presents the results for univariate regressions of excess market returns on each of the lagged highly persistent variables we consider in this study. We split the whole sample period into two halves. Panel A contains the results for the sub-period from February 1927 to June 1967, while Panel B contains the results for the sub-period from June 1967 to December 2008. When regressor *csp* is employed, Panel A contains results from July 1937 to June 1967 and Panel B includes results from July 1967 to December 2002. The general conclusion one can derive by inspecting the reported results is that there is no significant evidence in favour of predictability for excess market returns during the second half of the examined period in the case of univariate regressions. For all of the variables that we could not reject the null hypothesis in the whole sample period (d/e, lty, tbl, dfy and csp), this conclusion carries through in both of the sub-periods we consider. Therefore, the inability of these variables to predict next month excess market returns is robust to the choice of the period of analysis and it cannot be solely attributed to parameter instability. It is only for the term spread that we find weak evidence, at the 10% level, in favour of predictability during the second sub-period.

Examining the results for the variables that were found to be statistically significant predictors during the whole sample period (d/y, d/p, b/m, e/p and ntis), there is a degree of ambiguity. With respect to the dividend yield, this is found to be a significant predictor in the first period only at a 10% level, while there is no evidence in favour of predictability during the second period. For the dividend-price ratio we cannot reject the null hypothesis in any of these two periods. This finding is in line with Campbell and Yogo (2006), who documented much weaker evidence in favour of predictability in their post-1952 sample. This argument is further strengthened by the fact that the estimated coefficients are much smaller in magnitude during the second sub-period relative to the first one. Given that these are the two most commonly used variables, this evidence casts doubt on their predictive ability, especially when using recent sample periods and it should be taken into account by researchers who employ them for conditional asset pricing tests and conditional performance evaluation. With respect to the book to market value ratio and net equity expansion, which were among the most reliable predictors in the full sample analysis, we find that they could significantly predict excess market returns only for our pre-1967 sample. No such evidence is reported for the post-1967 sample. Overall our univariate regression results confirm the arguments of Lettau and van Nieuwerburgh (2008) regarding the notorious parameter instability due to the time-varying nature of this relationship and provide less support to strong conclusions derived using very long time series regressions.

At this point, it is very interesting to test the instability of inference with respect to the sub-periods investigated above, in the context of multivariate regressions. Table 4.9 presents the results for multivariate regressions of excess market returns on combinations of a subset of the regressors. The first line of each case reports the coefficient estimates of the regressors and the Wald statistic for overall significance for Panel A, while the second line reports the same quantities for Panel B. Due to limitations regarding the data mentioned above (lack of data of csp for the whole sample period, and interaction between dividend yield and dividend -price ratio, and interaction among the T-bill rate, long-term yield and term spread) we use only eight of the variables as regressors. All eight regressors are found to be insignificant in the context of univariate regressions for Panel B, apart from the term spread which exhibits weak evidence of significance (rejection of the null hypothesis of no significance at the 10% level). However, Table 4.9 shows that the same regressors appear to be jointly highly significant in the context of a multivariate regression. More specifically, the Wald statistic for overall significance of the regressors is 23.60739, resulting rejection of the no predictability hypothesis at 1%. At the second case (rows 3 and 4) of Table 4.9 the only regressor found to be weakly significant (tms), in the context of univariate regressions for Panel B, is dropped. A test of overall significance of the remaining regressors finds them to be jointly significant (rejection of the null hypothesis of no predictability at the 1% level). The difference in the conclusions drawn from univariate regressions and a multivariate regression (for the same set of regressors) is impressive. The Wald statistic, in the context of univariate regressions, cannot reject (not even at the 10% level of significance) the null of no predictability for any of the variables d/e, d/y, tbl, e/p, b/m dfy and ntis. In contradiction, when the joint significance of the aforementioned variables is tested, the Wald statistic strongly (at the 1% level) rejects the null hypothesis of no predictability. The third case of the same table refers to an equation including d/e, d/y, e/p, b/m dfy and ntis as regressors. In this case the regressors are found to be jointly insignificant for Panel B and jointly significant for Panel A. We then move to test the overall significance of the regressors found to be significant in the context of univariate regressions for the full sample period. These are variables d/y, e/p, b/m, dfy, and ntis. A test for their joint significance results to the rejection of the null hypothesis of no predictability for Panel A and no rejection of the same hypothesis for Panel B. Univariate regression analysis suggests insignificance of regressors d/e, tbl, dfy and tms for Panels A and B (with the exception of tms being significant at 10% level). A test of joint significance for these regressors suggests their significance at the 5% level for Panels A, while insignificance cannot be rejected for Panel B.

Comparison of the results presented in Tables 4.8 and 4.9 highlights the fact that a joint hypothesis test can lead to substantially different conclusion than the one resulting from the respective individual hypothesis tests. A characteristic example is testing the hypothesis of predictability using data from Panel B: individual hypothesis tests suggest that there is no predictability of stock returns, while a joint test of significance leads to the opposite answer. Table 4.9 shows that the magnitude of the Wald statistic is always higher, for each combination of regressors, for Panel A in comparison to Panel B. This could be considered as evidence of stronger predictability in the 1st period of the dataset examined.

4.5.2. Sensitivity to parameter choice

The results reported above are derived for a specific combination of the parameters β and γ that characterize the degree of persistence of the constructed instrumental variable and length of the truncation lag. The initial choice was to set $\beta = 5/6$ and $\gamma = 1/3$. The values of these parameters affect the derived Wald statistic, so it is legitimate to ask how this behaves for different combinations as well as if and how inference is modified relative to the benchmark case we have analyzed. To this end, we consider 170 combinations of these parameter values for the entire admissible set (i.e. $(\beta, \gamma) \in [2/3, 1) \times [0.25, 0.35]$). More specifically, we consider values of $\beta = 0.67, 0.69, ..., 0.99$, and $\gamma = 0.25, 0.26, ..., 0.35$. To visualize the sensitivity of the Wald statistic to the choice of parameter values, we plot the implied 3D surface along with a hyperplane that corresponds to the critical value at the 5% level for the degrees of freedom characterizing the examined case.

We perform this sensitivity analysis for each of the 11 variables employed to predict excess market returns in univariate regressions. With respect to the variables for which the null hypothesis of no predictability could not be rejected in the benchmark case (i.e. d/e, tbl, csp, dfy and tms), this conclusion is very robust to the combination of the parameter values we use. As an example of the variables in this category we plot in Figure 4.1 the corresponding 3D surface for the case when the T-bill rate is used as a regressor. It is evident that the Wald statistic (red surface) is not very sensitive to either β or γ values and it is always below the 5% critical value (blue hyperplane). With respect to the dividend-yield and the dividend-price ratio, this sensitivity analysis weakens further the evidence in favour of their individual predictive ability. Figure 4.2 plots the corresponding surface for the dividend yield. It is obvious that there are combinations of parameter values for which the null hypothesis of no predictability can be rejected, while for other combinations this is not true since the black hyperplane corresponding to the 5% critical value cuts through the coloured surface. This is particularly true for low values of γ as well as for very high or very low values of β. On the other hand, the earnings-price ratio is found to be a robust predictor of excess market returns for the full sample period. The most robust evidence in favour of predictability is provided by the book to market value ratio and net equity expansion. For any combination of these parameters' values, the null hypothesis of no predictability can be reliably rejected at levels even lower than 5%. For example, we plot in Figure 4.3 the generated surface for the book to market value ratio; this is well above the hyperplane corresponding to the critical value at the 5% level, robustifying the previous evidence in favour of predictability through this variable.

The overall conclusion with respect to the proposed methodology is that the Wald statistic does not seem to follow any monotonic pattern with respect to β and γ values, but its behaviour is case-specific. With respect to inference on predictability, we confirm that the borderline cases (i.e. cases where we marginally reject or fail to reject the null hypothesis at a specific confidence level) are subject to further ambiguity because the value of the Wald statistic depends on these parameters indeed. However, it should be stressed that the inference regarding marginal cases is always a notorious problem for any flexible econometric methodology that requires the use to choose parameter values. On the positive side, inference is not very sensitive to the choice of parameter values for cases where we reliably reject the null hypothesis of no predictability.

4.6. Conclusion

In this Chapter we employ the IVX methodology to the problem of testing the hypothesis of stock return predictability. The methodology built by MP and PM is extended by the inclusion of an intercept term in the predictive regression. This generalisation is motivated by the needs of applied work, as practitioners almost always include an intercept in the predictive regression in empirical studies. We apply this new methodology on data series that have been previously investigated as potential predictors of the market portfolio.

The empirical part of this Chapter describes the main conclusions drawn by the use of IVX methodology. The first is that the lagged series of the dividend price ratio, the dividend yield, the earnings price ratio, the book to market ratio and the net equity expansion appear to be significant for the determination of the market portfolio in the context of individual tests of significance and in the context of a joint test of overall significance. This can be considered as strong evidence of stock return predictability. The rest of the variables examined appear to be (individually and jointly) insignificant for the full sample period. The signs of the coefficients given by the IVX estimator are found to be compatible with both finance theory and previous empirical studies. Additionally, we investigate the answers provided by the IVX methodology when decomposed portfolios are used as explanatory variables. The results are interesting and suggest that there is a strong pattern linking predictability and the size of the portfolios. More specifically, predictability appears to be more evident for smaller market portfolios rather than larger ones. For portfolios ordered with respect to book to market value we find that, in the context of univariate regressions, predictability is in general present more often for high book to market portfolios. Book to market ratio and net equity expansion appear to be significant predictors for the aforementioned portfolios. However, joint hypothesis tests suggest that default yield spread is the only significant predictor for all portfolios decomposed according to their book to market value.

The inference drawn by the IVX methodology is also examined for two subperiods of the available sample size. Using univariate regressions we find that predictability of stock returns exists in the first period but vanishes in the second period. Interestingly, using multivariate regressions (i.e. testing the joint significance of the regressors) provides a different answer with respect to predictability in the second period: variables that are individually insignificant appear to be jointly significant (even at the 1% level).

The above observation is only one of numerous examples discovered throughout this study where conclusions based on joint inference on a multiple system of predictive regressions may differ from those drawn from individual hypothesis tests. This highlights one of the advantages of IVX methodology over existing methods based on local to unity univariate regressions. The ability of the IVX method to accommodate joint testing of general linear restrictions on the predictive variables can be a valuable tool for practitioners, as it extends the range of testable hypotheses and models and can provide different answers on more sophisticated empirical problems than individual tests of significance.

Another appealing feature of IVX inference is its robustness to various time series modelling frameworks including unit root, local to unity and moderate to unity persistence structure. Robustness of the method to the degree of regressor persistence is crucial given the fact that the parameters c_i and a cannot be jointly estimated and misspecification can lead to seriously distorted inference.

4.7. Technical appendix and proofs

This Section contains the proof Lemma 4.1 and Theorem 4.1. We begin by establishing some technical lemmata that facilitate the above proofs.

Lemma 7.1. *Let u_t be a linear process given by (4.5) satisfying (4.6) and $E\left\|\varepsilon_1\right\|^4 < \infty$, and let $\Gamma_u(h) = E\left(u_t u_{t-h}'\right)$. Then, there exists $B > 0$ such that*

$$\max_{h \leq n} E\left\|\frac{1}{n^{1/2}}\sum_{t=h+1}^{n}\left[u_t u_{t-h}' - \Gamma_u(h)\right]\right\| \leq B < \infty.$$

Lemma 7.1 can be proved by an identical argument to that used in the proof of Proposition A2 in MP.

Lemma 7.2. *Let $\alpha > 1/2$ and $M = L(n)\, n^{\gamma}$ with $\gamma \in (0, 1/2)$. The following orders of magnitude apply uniformly for any $h \in \{1, ..., M\}$:*

(i) $\sum_{t=h}^{n} u_{0t} x_{t-h}' = O_p(n)$.

(ii) $\sum_{t=h+1}^{n} x_t u_{xt-h}' = O_p(n)$.

(iii) $\sum_{t=h+1}^{n} x_t x_{t-h-1}' = O_p\left(n^{1+\alpha}\right)$.

Proof. For part (i), the BN decomposition and summation by parts yield

$$\begin{aligned}
\sum_{t=h+1}^{n} u_{0t} x_{t-h-1}' &= F_0(1)\sum_{t=h+1}^{n}\varepsilon_t x_{t-h-1}' - \sum_{t=h+1}^{n}\Delta\tilde{\varepsilon}_{0t} x_{t-h-1}' \\
&= -\sum_{t=h+1}^{n}\Delta\tilde{\varepsilon}_{0t} x_{t-h-1}' + O_p\left(n^{(1+\alpha)/2}\right) \\
&= \sum_{t=h+1}^{n}\tilde{\varepsilon}_{0t}\Delta x_{t-h}' + O_p\left(n^{(1+\alpha)/2}\right)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{n^{\alpha}} \sum_{t=h+1}^{n} \tilde{\varepsilon}_{0t} x_{t-h-1}' C + \sum_{t=h+1}^{n} \tilde{\varepsilon}_{0t} u_{xt-h}' + O_p\left(n^{(1+\alpha)/2}\right) \\
&= \frac{1}{n^{\alpha}} \sum_{t=h+1}^{n} \tilde{\varepsilon}_{0t} x_{t-h-1}' C + \sum_{t=1}^{n} \tilde{\varepsilon}_{0t} u_{xt-h}' + O_p(M) + O_p\left(n^{(1+\alpha)/2}\right) \\
&= O_p(n)
\end{aligned}
$$

by the ergodic theorem since

$$
\begin{aligned}
E\left\| \frac{1}{n^{1+\alpha}} \sum_{t=h+1}^{n} \left(x_{t-h-1} \otimes \tilde{\varepsilon}_{0t}\right) \right\| &\leq \frac{1}{n^{1+\alpha}} \sum_{t=h+1}^{n} E\left(\|x_{t-h-1}\| \, \|\tilde{\varepsilon}_{0t}\|\right) \\
&\leq \frac{1}{n^{1+\alpha}} \sum_{t=1}^{n} \left(E\|x_{t-h-1}\|^2\right)^{1/2} \left(E\|\tilde{\varepsilon}_{0t}\|^2\right)^{1/2} \\
&= O\left(\frac{n^{1+\alpha/2}}{n^{1+\alpha}}\right) = O\left(\frac{1}{n^{\alpha/2}}\right).
\end{aligned}
$$

For part (ii), write

$$
\begin{aligned}
\sum_{t=h+1}^{n} x_t u_{xt-h}' &= \sum_{t=h+1}^{n} x_{t-h-1} u_{xt-h}' + \sum_{t=h+1}^{n} \left(x_t - x_{t-h-1}\right) u_{xt-h}' \\
&= \sum_{t=1}^{n-h} x_{t-1} u_{xt}' + \sum_{t=h+1}^{n} \left(x_t - x_{t-h-1}\right) u_{xt-h}' \\
&= \sum_{t=h+1}^{n} \left(x_t - x_{t-h-1}\right) u_{xt-h}' + O_p(n) \qquad (4.30)
\end{aligned}
$$

by an identical argument to part (i). Recall that $x_t = R_n^t x_0 + \sum_{j=1}^t R_n^{t-j} u_{xj}$. So, ignoring for the moment the initial condition, $x_t - x_{t-h-1}$ can be written as

$$\sum_{j=1}^{t} R_n^{t-j} u_{xj} - \sum_{j=1}^{t-h-1} R_n^{t-h-1-j} u_{xj} = \sum_{j=t-h}^{t} R_n^{t-j} u_{xj} + \left(I_K - R_n^{-h-1}\right) \sum_{j=1}^{t-h-1} R_n^{t-j} u_{xj} \tag{4.31}$$

and note that, since $h \leq M \ll n^\alpha$,

$$\begin{aligned}
I_K - R_n^{-h-1} &= R_n^{-h-1}\left(R_n^{h+1} - I_K\right) \\
&= R_n^{-h-1}\left[\exp\left\{(h+1)\log\left(I_K + C/n^\alpha\right)\right\} - I_K\right] \\
&= R_n^{-h-1}\left[\exp\left\{(h+1)\left[\frac{C}{n^\alpha} + O\left(\frac{1}{n^{2\alpha}}\right)\right]\right\} - I_K\right] \\
&= R_n^{-h-1}\left[e^{\frac{M}{n^\alpha}C + O\left(\frac{M}{n^{2\alpha}}\right)} - I_K\right] \\
&= R_n^{-h-1}\frac{M}{n^\alpha}C + O\left(\frac{M^2}{n^{2\alpha}}\right) \\
&= \left[I_K + O\left(\frac{M}{n^\alpha}\right)\right]\frac{M}{n^\alpha}C + O\left(\frac{M^2}{n^{2\alpha}}\right) \\
&= \frac{M}{n^\alpha}C + O\left(\frac{M^2}{n^{2\alpha}}\right).
\end{aligned} \tag{4.32}$$

Combining (4.31) and (4.32) we obtain

$$\begin{aligned}
\sum_{t=h+1}^{n}\left(x_t - x_{t-h-1}\right)u'_{xt-h} &= \sum_{t=h+1}^{n}\left(\sum_{j=t-h}^{t} R_n^{t-j}u_{xj}\right)u'_{xt-h} \\
&\quad + \frac{M}{n^{\alpha}}C\sum_{t=h+1}^{n}\left(\sum_{j=1}^{t-h-1} R_n^{t-j}u_{xj}\right)u'_{xt-h} \\
&\quad + \left(I_K - R_n^{-h-1}\right)\sum_{t=h+1}^{n} R_n^{t}x_0u'_{xt-h}. \qquad (4.33)
\end{aligned}$$

By (4.32) and a standard CLT, the last term in (4.33) has order

$$O_p\left(Mn^{-\alpha}x_0n^{\alpha/2}\right) = o_p\left(M\right).$$

For the first term of (4.33), letting $k = t - j$, we obtain

$$\begin{aligned}
\sum_{t=h+1}^{n}\left(\sum_{j=t-h}^{t} R_n^{t-j}u_{xj}\right)u'_{xt-h} &= \sum_{t=h+1}^{n}\sum_{k=0}^{h} R_n^{k}u_{xt-k}u'_{xt-h} \\
&= \sum_{k=0}^{h} R_n^{k}\sum_{t=h+1}^{n}\left[u_{xt-k}u'_{xt-h} - \Gamma_{xx}\left(h-k\right)\right] \\
&\quad + \sum_{k=0}^{h} R_n^{k}\left(n-h\right)\Gamma_{xx}\left(h-k\right) \\
&= O_p\left(n\right)
\end{aligned}$$

because the second term is bounded by $n\sum_{k=0}^{\infty}\|\Gamma_{xx}(k)\|$ and, letting $i=t-h$, the first term is bounded in L_1 norm by

$$\begin{aligned}
& \max_{h,k} E\left\|\sum_{t=1}^{n-h}\left[u_{xi+h-k}u'_{xi}-\Gamma_{xx}(h-k)\right]\right\|\sum_{k=0}^{h}\left\|R_n^k\right\| \\
\leq\ & M\max_{0\leq l\leq h} E\left\|\sum_{t=1}^{n-h}\left[u_{xi+l}u'_{xi}-\Gamma_{xx}(l)\right]\right\| \\
=\ & O\left(Mn^{1/2}\right)=o(n)
\end{aligned}$$

for any $\gamma<1/2$ by Lemma 7.1.

The second term of (4.33) can be dealt with in the usual way since the "regressor" belongs to the past of the "innovation": using the BN decomposition and summation by parts

$$\begin{aligned}
& \frac{M}{n^{\alpha}}\sum_{t=h+1}^{n}\left(\sum_{j=1}^{t-h-1}R_n^{t-j}u_{xj}\right)u'_{xt-h} \\
=\ & \frac{M}{n^{\alpha}}\sum_{t=h+1}^{n}\left(\sum_{j=1}^{t-h-1}R_n^{t-j}u_{xj}\right)\varepsilon'_{t-h}F_x(1)'-\frac{M}{n^{\alpha}}\sum_{t=h+1}^{n}\left(\sum_{j=1}^{t-h-1}R_n^{t-j}u_{xj}\right)\Delta\tilde{\varepsilon}'_{xt-h} \\
=\ & -\frac{M}{n^{\alpha}}\left\{O_p\left(\frac{1}{n^{\alpha/2}}\right)-\sum_{t=h+1}^{n}\left(\sum_{j=1}^{t-h}R_n^{t+1-j}u_{xj}-\sum_{j=1}^{t-h-1}R_n^{t-j}u_{xj}\right)\tilde{\varepsilon}'_{xt-h}\right\} \\
& +O_p\left(\frac{Mn^{\frac{1+\alpha}{2}}}{n^{\alpha}}\right)
\end{aligned}$$

$$
\begin{aligned}
&= R_n^{h+1}\frac{M}{n^{\alpha}}\sum_{t=h+1}^{n}u_{xt-h-1}\tilde{\varepsilon}'_{xt-h}+\frac{M}{n^{\alpha}}\left(R_n-I_K\right)\sum_{t=h+1}^{n}\sum_{j=1}^{t-h-1}R_n^{t-j}u_{xj}\tilde{\varepsilon}'_{xt-h}\\
&\quad +O_p\left(\frac{Mn^{1/2}}{n^{\alpha/2}}\right)\\
&= O_p\left(\frac{Mn}{n^{\alpha}}\right)+C\frac{M}{n^{2\alpha}}\sum_{t=h+1}^{n}\left(\sum_{j=1}^{t-h-1}R_n^{t-j}u_{xj}\right)\tilde{\varepsilon}'_{xt-h}\\
&= O_p\left(\frac{Mn}{n^{\alpha}}\right)=o_p\left(n\right)
\end{aligned}
$$

for any $\alpha > 1/2$ and $\gamma < 1/2$. The result follows by (4.30) and (4.33).

For part (iii), using the recursive property of x_t we obtain

$$
x_t x'_{t-h-1} = R_n x_{t-1}x'_{t-h-2}R_n + R_n x_{t-1}u'_{xt-h-1} + u_{xt}x'_{t-h-2}R_n + u_{xt}u'_{xt-h-1},
$$

so summing and using (i), (ii) and the LLN we obtain

$$
\frac{1}{n^{\alpha}}\sum_{t=h+1}^{n}x_t x'_{t-h-1} = O_p\left(n\right).
$$

Lemma 7.3.

(i) *Partitioning the OLS estimator in (4.20)* $\underline{\tilde{A}}_{OLS} = \left[\tilde{\mu}_{OLS}, \tilde{A}_{OLS}\right]$ *conformably to* $\underline{A} = [\mu, A]$ *the OLS estimators of* μ *and* A *are given by*

$$
\tilde{\mu}_{OLS} = \frac{1}{\phi_n}\left(\sum_{t=1}^{n}y_t - \hat{A}_{OLS}\sum_{t=1}^{n}x_{t-1}\right) \tag{4.34}
$$

and

$$\tilde{A}_{OLS} = \hat{A}_{OLS} - \tilde{\mu}_{OLS} \left(\sum_{t=1}^{n} x_{t-1} \right)' \left(\sum_{t=1}^{n} x_{t-1} x_{t-1}' \right)^{-1}, \tag{4.35}$$

where

$$\phi_n = n - \left(\sum_{t=1}^{n} x_{t-1} \right)' \left(\sum_{t=1}^{n} x_{t-1} x_{t-1}' \right)^{-1} \left(\sum_{t=1}^{n} x_{t-1} \right) \tag{4.36}$$

and $\hat{A}_{OLS} = \left(\sum_{t=1}^{n} y_t x_{t-1}' \right) \left(\sum_{t=1}^{n} x_{t-1} x_{t-1}' \right)^{-1}$ *is the OLS estimator when* $\mu = 0$.

(ii) *For any* $\alpha > 0$ *the OLS estimators in (4.35) and (4.34) have consistency rates*

$$\begin{aligned} \tilde{A}_{OLS} - A &= O_p \left(n^{-(\alpha \wedge 1)} \right) \\ \tilde{\mu}_{OLS} - \mu &= O_p \left(n^{-1/2} \right). \end{aligned}$$

Proof. Part (i) of the lemma is a consequence of a standard partitioned inverse formula, see 5.29 in Abadir and Magnus (2005).

For part (ii), note that ϕ_n has exact order of magnitude equal to $O_e(n)$ for all $\alpha > 0$ since

$$\frac{\phi_n}{n} = 1 - \frac{1}{n^{1-\alpha}} \left(\frac{1}{n^{\frac{1}{2}+\alpha}} \sum_{t=1}^{n} x_{t-1} \right)' \left(\frac{1}{n^{1+\alpha}} \sum_{t=1}^{n} x_{t-1} x_{t-1}' \right)^{-1} \left(\frac{1}{n^{\frac{1}{2}+\alpha}} \sum_{t=1}^{n} x_{t-1} \right),$$

so, if $\alpha = 1$,

$$\frac{\phi_n}{n} \Rightarrow 1 - \left(\int_0^1 J_C ds\right)' \left(\int_0^1 J_C J_C' ds\right)^{-1} \left(\int_0^1 J_C ds\right), \tag{4.37}$$

and $\phi_n/n \to_p 1$ if $\alpha \in (0,1)$. As usual, the limit of ϕ_n/n unit root case may be obtained by setting $C = 0$ in the local to unity case which amounts to replacing J_C by B_x in (4.37).

Let

$$\Psi_n = \left(\sum_{t=1}^n u_{0t} x_{t-1}'\right) \left(\sum_{t=1}^n x_{t-1} x_{t-1}'\right)^{-1}.$$

Of course, Ψ_n is not affected by the presence of an intercept and $\Psi_n = O_p\left(n^{-\alpha}\right)$ as before. Using the identities

$$\sum_{t=1}^n y_t x_{t-1}' = \mu \left(\sum_{t=1}^n x_{t-1}\right)' + A \sum_{t=1}^n x_{t-1} x_{t-1}' + \sum_{t=1}^n u_{0t} x_{t-1}' \tag{4.38}$$

and

$$\sum_{t=1}^n y_t = n\mu + A \sum_{t=1}^n x_{t-1} + \sum_{t=1}^n u_{0t}$$

we obtain

$$\begin{aligned}
\tilde{\mu}_{OLS} &= \frac{1}{\phi_n} \left[\sum_{t=1}^n y_t - \left(\sum_{t=1}^n y_t x_{t-1}'\right) \left(\sum_{t=1}^n x_{t-1} x_{t-1}'\right)^{-1} \sum_{t=1}^n x_{t-1}\right] \\
&= \frac{1}{\phi_n} \left[n\mu + A \sum_{t=1}^n x_{t-1} + \sum_{t=1}^n u_{0t} - \left(\mu\left(n - \phi_n\right) + A \sum_{t=1}^n x_{t-1} + \Psi_n \sum_{t=1}^n x_{t-1}\right)\right]
\end{aligned}$$

$$
\begin{aligned}
&= \mu + \frac{1}{\phi_n}\left(\sum_{t=1}^{n} u_{0t} - \Psi_n \sum_{t=1}^{n} x_{t-1}\right) \\
&= \mu + O_p\left(n^{-1/2}\right)
\end{aligned}
$$

by the CLT, since $\phi_n = O_e(n)$ and $\Psi_n \sum_{t=1}^{n} x_t = O_p\left(n^{-\alpha} n^{1/2+\alpha}\right) = O_p\left(n^{1/2}\right)$.

For $\tilde{A}_{OLS}$ first note that, using (4.38), we obtain

$$
\begin{aligned}
\hat{A}_{OLS} &= \left(\sum_{t=1}^{n} y_t x_{t-1}'\right)\left(\sum_{t=1}^{n} x_{t-1} x_{t-1}'\right)^{-1} \\
&= A + \mu\left(\sum_{t=1}^{n} x_{t-1}\right)'\left(\sum_{t=1}^{n} x_{t-1} x_{t-1}'\right)^{-1} + \Psi_n.
\end{aligned}
$$

Substituting into (4.35) we get

$$
\begin{aligned}
\tilde{A}_{OLS} - A &= \Psi_n - \left(\tilde{\mu}_{OLS} - \mu\right)\left(\sum_{t=1}^{n} x_{t-1}\right)'\left(\sum_{t=1}^{n} x_{t-1} x_{t-1}'\right)^{-1} \\
&= \Psi_n + O_p\left(n^{-1}\right) \\
&= O_p\left(n^{-\alpha}\right)
\end{aligned}
$$

as required.

Proof of Lemma 4.1 (i). Using the identities

$$
\tilde{u}_{0t} = u_{0t} - \left(\tilde{A}_{OLS} - A\right) x_{t-1} - \left(\tilde{\mu}_{OLS} - \mu\right). \tag{4.39}
$$

and

$$
\hat{u}_{xt-h} = u_{xt-h} - \left(\hat{R}_n - R_n\right) x_{t-h-1}
$$

we obtain the following expansion of (4.23):

$$\hat{\Lambda}_{0x}^{(n)} = \hat{\Lambda}_1^{(n)} - \hat{\Lambda}_2^{(n)} - \hat{\Lambda}_3^{(n)} + \hat{\Lambda}_4^{(n)} - \hat{\Lambda}_4^{(5)} + \hat{\Lambda}_4^{(6)} \tag{4.40}$$

where

$$\begin{aligned}
\hat{\Lambda}_1^{(n)} &= \frac{1}{n}\sum_{h=1}^{M}\left(1-\frac{h}{M+1}\right)\sum_{t=h+1}^{n} u_{0t}u'_{xt-h} \\
\hat{\Lambda}_2^{(n)} &= \frac{1}{n}\sum_{h=1}^{M}\left(1-\frac{h}{M+1}\right)\sum_{t=h+1}^{n} u_{0t}x'_{t-h-1}\left(\hat{R}_n - R_n\right)' \\
\hat{\Lambda}_3^{(n)} &= \left(\tilde{A}_{OLS} - A\right)\frac{1}{n}\sum_{h=1}^{M}\left(1-\frac{h}{M+1}\right)\sum_{t=h+1}^{n} x_t u'_{xt-h} \\
\hat{\Lambda}_4^{(n)} &= \left(\tilde{A}_{OLS} - A\right)\frac{1}{n}\left[\sum_{h=1}^{M}\left(1-\frac{h}{M+1}\right)\sum_{t=h+1}^{n} x_t x'_{t-h-1}\right]\left(\hat{R}_n - R_n\right)' \\
\hat{\Lambda}_5^{(n)} &= \left(\tilde{\mu}_{OLS} - \mu\right)\frac{1}{n}\sum_{h=1}^{M}\left(1-\frac{h}{M+1}\right)\sum_{t=h+1}^{n} u'_{xt-h} \\
\hat{\Lambda}_6^{(n)} &= \left(\tilde{\mu}_{OLS} - \mu\right)\frac{1}{n}\sum_{h=1}^{M}\left(1-\frac{h}{M+1}\right)\sum_{t=h+1}^{n} x'_{t-h-1}\left(\hat{R}_n - R_n\right).
\end{aligned}$$

Using Lemma 7.2, Lemma 7.3 and the fact that, by equation (11) of MP, $\hat{R}_n - R_n = O_p\left(n^{-\alpha}\right)$ we obtain that

$$\hat{\Lambda}_{0x}^{(n)} = \hat{\Lambda}_1^{(n)} + O_p\left(\frac{M}{n^{\alpha}}\right). \tag{4.41}$$

Therefore, since $\alpha > 1/2$, establishing

$$\hat{\Lambda}_1^{(n)} = \Lambda_{0x} + O_p\left(\frac{M}{n^{1/2}}\right) + O_p\left(\frac{1}{M}\right) \tag{4.42}$$

is sufficient to for the proof of the lemma. To prove (4.42), letting

$$\Gamma_{0x}(h) = E\left(u_{0t}u'_{xt-h}\right)$$

we can write

$$\begin{aligned} \hat{\Lambda}_1^{(n)} &= \frac{1}{n}\sum_{h=1}^{M}\left(1-\frac{h}{M+1}\right)\left[u_{0t}u'_{xt-h} - \Gamma_{0x}(h)\right] \\ &\quad + \frac{1}{n}\sum_{h=1}^{M}\left(1-\frac{h}{M+1}\right)(n-h)\,\Gamma_{0x}(h). \end{aligned} \tag{4.43}$$

The first term of (4.43) has order $O_p\left(\frac{M}{n^{1/2}}\right)$ since it is bounded in L_1 norm by

$$\frac{1}{n}\max_{h\leq M} E\left\|\sum_{t=h+1}^{n}\left[u_{0t}u'_{xt-h} - \Gamma_{0x}(h)\right]\right\| M = O\left(\frac{M}{n^{1/2}}\right)$$

by Lemma 7.1. The summability assumption (4.6) implies that

$$\sum_{h=1}^{\infty} h\left\|\Gamma_{0x}(h)\right\| < \infty. \tag{4.44}$$

Using (4.44), the second term of (4.43) can be written as

$$\begin{aligned}
\sum_{h=1}^{M}\left(1-\frac{h}{M+1}\right)\left(1-\frac{h}{n}\right)\Gamma_{0x}(h) &= \sum_{h=1}^{M}\Gamma_{0x}(h)+\sum_{h=1}^{M}\frac{h}{M+1}\Gamma_{0x}(h)+O\left(\frac{M}{n}\right) \\
&= \sum_{h=1}^{M}\Gamma_{0x}(h)+O\left(\frac{1}{M}\right) \\
&= \Lambda_{0x}+O\left(\frac{1}{M}\right)
\end{aligned}$$

because

$$\begin{aligned}
\left\|\Lambda_{0x}-\sum_{h=1}^{M}\Gamma_{0x}(h)\right\| &= \left\|\sum_{h=M+1}^{\infty}\Gamma_{0x}(h)\right\| \leq \sum_{h=M+1}^{\infty}\left\|\Gamma_{0x}(h)\right\| \\
&\leq \frac{1}{M}\sum_{h=M+1}^{\infty}h\left\|\Gamma_{0x}(h)\right\| = O\left(\frac{1}{M}\right).
\end{aligned}$$

This shows (4.42) and the lemma.

Proof of Lemma 4.1 (ii). Set $\beta\in(2/3,1)$ and $M=L(n)\,n^{\gamma}$ for $\gamma\geq 1/4$. We distinguish between the cases $\alpha>\beta$ and $\alpha\leq\beta$.

When $\alpha>\beta$, part (i) of the lemma yields

$$\begin{aligned}
n^{\frac{1-(\alpha\wedge\beta)}{2}}\left(\hat{\Lambda}_{0x}-\Lambda_{0x}\right) &= n^{\frac{1-\beta}{2}}\left(\hat{\Lambda}_{0x}-\Lambda_{0x}\right) \\
&= O_p\left(\frac{M}{n^{\beta/2}}\right)+O_p\left(\frac{n^{\frac{1-\beta}{2}}}{M}\right) \\
&= O_p\left(\frac{M}{n^{\beta/2}}\right)=O_p\left(\frac{L(n)\,n^{\gamma}}{n^{\beta/2}}\right)
\end{aligned}$$

because $M/n^{\beta/2} \gg n^{\frac{1-\beta}{2}}/M$ for any $\gamma \geq 1/4$. Since slowly varying functions increase to infinity slower than any polynomial, the above order of magnitude will tend to 0 if and only if $\beta/2 > \gamma$. So, any choice of β in $(2/3, 1)$ achieves the optimal bandwidth selection $\gamma = 1/3$.

When $\alpha \leq \beta$, an identical calculation yields

$$\begin{aligned} n^{\frac{1-(\alpha\wedge\beta)}{2}}\left(\hat{\Lambda}_{0x} - \Lambda_{0x}\right) &= n^{\frac{1-\alpha}{2}}\left(\hat{\Lambda}_{0x} - \Lambda_{0x}\right) \\ &= O_p\left(\frac{M}{n^{\alpha/2}}\right) = O_p\left(\frac{L(n)\, n^{\gamma}}{n^{\alpha/2}}\right). \end{aligned}$$

Therefore, in order for condition (4.24) to be satisfied for all $\alpha > 1/2$, the bandwidth choice is restricted to $\gamma = 1/4$. The optimal bandwidth selection $\gamma = 1/3$ only applies if we impose the additional restriction $\alpha > 2/3$.

Most of the sample moment limit theory needed for the proof of Theorem 4.1 can be found in the original papers by MP and PM. The next lemma discusses the asymptotic behaviour of the sample mean of the IVX instruments in (4.12) that arises as a result of including an intercept in (4.2).

Lemma 7.4. *The following approximations are valid as $n \to \infty$:*

(i) When $\beta < \min(\alpha, 1)$:

$$\frac{1}{n^{\frac{1}{2}+\beta}} \sum_{t=1}^{n} \tilde{z}_t = -C_z^{-1}\left(\frac{1}{n^{1/2}} \sum_{t=1}^{n} u_{xt} + \frac{C}{n^{1/2+\alpha}} \sum_{t=1}^{n} x_{t-1}\right) + O_p\left(n^{-(1-\beta)/2}\right). \quad (4.45)$$

(ii) When $1/3 < \alpha \leq \beta < 1$:

$$\frac{1}{n^{\frac{1}{2}+\alpha}} \sum_{t=1}^{n} \tilde{z}_t \rightarrow_p 0. \tag{4.46}$$

Proof. For part (i), using the decomposition $\tilde{z}_t = z_t + \frac{C}{n^\alpha}\psi_{nt}$ we obtain

$$\frac{1}{n^{\frac{1}{2}+\beta}} \sum_{t=1}^{n} \tilde{z}_t = \frac{1}{n^{\frac{1}{2}+\beta}} \sum_{t=1}^{n} z_t + \frac{C}{n^{\frac{1}{2}+\beta+\alpha}} \sum_{t=1}^{n} \psi_{nt}. \tag{4.47}$$

For the first term of (4.47), summing $z_t = R_{nz} z_{t-1} + u_{xt}$ yields

$$(I_K - R_{nz}) \sum_{t=1}^{n} z_t = \sum_{t=1}^{n} u_{xt} + O_p(z_n).$$

Since $I_K - R_{nz} = -C_z/n^\beta$ and $z_n = O_p\left(n^{\beta/2}\right)$ we obtain

$$\frac{1}{n^{\frac{1}{2}+\beta}} \sum_{t=1}^{n} z_t = -C_z^{-1} \frac{1}{n^{1/2}} \sum_{t=1}^{n} u_{xt} + O_p\left(n^{-(1-\beta)/2}\right). \tag{4.48}$$

For the second term of (4.47), summing the recursive formula (see equation (44) of PM)

$$\psi_{nt} = R_{nz}\psi_{n,t-1} + x_{t-1}.$$

we obtain

$$\frac{1}{n^{1/2+\alpha+\beta}} \sum_{t=1}^{n} \psi_{nt} = -C_z^{-1} \frac{1}{n^{1/2+\alpha}} \sum_{t=1}^{n} x_{t-1} + O_p\left(\frac{\psi_{n,n}}{n^{1/2+\alpha}}\right). \tag{4.49}$$

By Proposition A2 of PM, $\psi_{n,n} = O_p\left(n^{\alpha/2+\beta}\right)$ for all $\beta < \alpha$ so

$$\frac{\psi_{n,n}}{n^{1/2+\alpha}} = O_p\left(\frac{n^{\beta}}{n^{\frac{1+\alpha}{2}}}\right) = o_p\left(n^{-\frac{1-\beta}{2}}\right).$$

Part (i) now follows by combining (4.47), (4.48) and (4.49).

For part (ii), using the decomposition

$$\tilde{z}_t = x_t - R_{nz}^t x_0 + \frac{C_z}{n^{\beta}}\psi_{nt},$$

see equation (23) in PM, we obtain

$$\frac{1}{n^{1/2+\alpha}}\sum_{t=1}^{n}\tilde{z}_t = \frac{1}{n^{1/2+\alpha}}\sum_{t=1}^{n}x_t + \frac{C_z}{n^{1/2+\alpha+\beta}}\sum_{t=1}^{n}\psi_{nt} + o_p(1).$$

Substituting (4.49) to the above display we obtain

$$\begin{aligned}
\frac{1}{n^{1/2+\alpha}}\sum_{t=1}^{n}\tilde{z}_t &= \frac{1}{n^{1/2+\alpha}}\sum_{t=1}^{n}x_t - \frac{1}{n^{1/2+\alpha}}\sum_{t=1}^{n}x_{t-1} + O_p\left(\frac{\psi_{n,n}}{n^{1/2+\alpha}}\right) \\
&= \frac{x_n}{n^{1/2+\alpha}} - \frac{x_0}{n^{1/2+\alpha}} + O_p\left(\frac{\psi_{n,n}}{n^{1/2+\alpha}}\right) \\
&= O_p\left(n^{-\frac{1-\beta}{2}}\right)
\end{aligned}$$

since $x_n = O_p\left(n^{\alpha/2}\right)$ and $\psi_{n,n} = O_p\left(n^{\alpha+\beta/2}\right)$ for all $\beta \geq \alpha$ by Proposition A2 of PM.

Proof of Theorem 4.1. We use Lemma 4.7 throughout.

For part (i), we start with the signal matrix:

$$\begin{aligned}
\underline{X}'\tilde{Z} &= X'\tilde{Z} - n\bar{x}_{n-1}\bar{\tilde{z}}_{n-1}' \\
&= X'\tilde{Z} - \frac{1}{n}\left(\sum_{t=1}^{n} x_{t-1}\right)\left(\sum_{t=1}^{n} \tilde{z}_{t-1}\right)'.
\end{aligned}$$

The limit distribution of $n^{-(1+\beta)}X'\tilde{Z}$ is given by Lemma 3.1(ii) and equation (20) of PM. Using (4.45) we obtain

$$\begin{aligned}
\frac{1}{n^{1+\beta}}\underline{X}'\tilde{Z} &= \frac{X'\tilde{Z}}{n^{1+\beta}} - \left(\frac{1}{n^{3/2}}\sum_{t=1}^{n} x_{t-1}\right)\left(\frac{1}{n^{\frac{1}{2}+\beta}}\sum_{t=1}^{n} \tilde{z}_{t-1}\right)' \\
&= \frac{X'\tilde{Z}}{n^{1+\beta}} + \left(\frac{1}{n^{3/2}}\sum_{t=1}^{n} x_{t-1}\right)\left(\frac{1}{n^{1/2}}\sum_{t=1}^{n} u_{xt} + \frac{C}{n^{1/2+\alpha}}\sum_{t=1}^{n} x_{t-1}\right)' C_z^{-1} \\
&\quad + o_p(1).
\end{aligned}$$

Note that all of the above normalised sums are bounded in probability for all $\alpha > 0$. When $\alpha = 1$ (x_t is a local to unity process),

$$\begin{aligned}
\frac{1}{n^{1+\beta}}\underline{X}'\tilde{Z} &\Rightarrow -\left[\int_0^1 J_C dB_x' + \Omega_{xx} + \int_0^1 J_C J_C' C\right] C_z^{-1} \\
&\quad + \left(\int_0^1 J_C\right)\left[B_x(1)' + \left(\int_0^1 J_C\right)' C\right] C_z^{-1} \\
&= -\left[\Omega_{xx} + \int_0^1 \underline{J}_C dB_x' + \int_0^1 \underline{J}_C \underline{J}_C' ds C\right] C_z^{-1} \qquad (4.50)
\end{aligned}$$

where $\underline{J}_C(t) = J_C(t) - \int_0^1 J_C(t)\,dt$ and $J_C(t) = \int_0^t e^{(t-s)C} dB_x(s)$. In the unit root case of P(i), the limit distribution of $n^{-(1+\beta)}\underline{X}'\tilde{Z}$ can be obtained by substituting

$C = 0$ in (4.50):

$$\frac{1}{n^{1+\beta}}\underline{X}'\tilde{Z} \Rightarrow -\left[\Omega_{xx} + \int_0^1 \underline{B}_x dB_x'\right] C_z^{-1}, \tag{4.51}$$

where $\underline{B}_x(t) = B_x(t) - \int_0^1 B_x(t)\,dt$. In the mildly integrated case, $\sum_{t=1}^n x_{t-1} = O_p\left(n^{1/2+\alpha}\right)$ with $\alpha < 1$, so $n^{-3/2}\sum_{t=1}^n x_{t-1} = o_p(1)$ giving

$$\frac{1}{n^{1+\beta}}\underline{X}'\tilde{Z} = \frac{X'\tilde{Z}}{n^{1+\beta}} + o_p(1) = -\left(\Omega_{xx} + V_{xx}C\right)C_z^{-1} + o_p(1) \tag{4.52}$$

by equation (7) of MP and Lemma 3.1(ii) and equation (20) of PM. Combining (4.50), (4.51) and (4.52) and taking into account multiplication by $-C_z^{-1}$ yields $\tilde{\Psi}_{xx}$ of Theorem 4.1.

Next, we show that the presence of an intercept in (4.2) has no effect on the asymptotic behaviour of the $\underline{U}_0'\tilde{Z}$ matrix:

$$\begin{aligned} n^{-(1+\beta)/2}\underline{U}_0'\tilde{Z} &= n^{-(1+\beta)/2}U_0'\tilde{Z} - n^{(1-\beta)/2}\bar{u}_{0,n}\bar{z}_{n-1}' \\ &= n^{-(1+\beta)/2}U_0'\tilde{Z} - \left(\frac{1}{n^{1/2}}\sum_{t=1}^n u_{0t}\right)\left(\frac{1}{n^{1+\beta/2}}\sum_{t=1}^n \tilde{z}_{t-1}\right)' \\ &= n^{-(1+\beta)/2}U_0'\tilde{Z} + O_p\left(n^{-\frac{1-\beta}{2}}\right) \end{aligned}$$

by (4.45) and the CLT.

When $\alpha \leq \beta < 1$ we show that the presence of an intercept in (4.2) has no effect on IVX limit theory. Since $n^{-3/2}\sum_{t=1}^n x_{t-1}$ and $n^{-(1+\alpha/2)}\sum_{t=1}^n \tilde{z}_t$ are both

$o_p(1)$, The signal matrix

$$\begin{aligned} n^{-(1+\alpha)}\underline{X}'\tilde{Z} &= n^{-(1+\alpha)}X'\tilde{Z} - n^{-\alpha}\bar{x}_{n-1}\bar{z}_{n-1}' \\ &= n^{-(1+\alpha)}X'\tilde{Z} - \left(\frac{1}{n^{3/2}}\sum_{t=1}^{n} x_{t-1}\right)\left(\frac{1}{n^{\frac{1}{2}+\alpha}}\sum_{t=1}^{n} \tilde{z}_{t-1}\right)' \\ &= n^{-(1+\alpha)}X'\tilde{Z} + o_p(1) \end{aligned}$$

and

$$\begin{aligned} n^{-(1+\alpha)/2}\underline{U}_0'\tilde{Z} &= n^{-(1+\alpha)/2}U_0'\tilde{Z} - n^{(1-\alpha)/2}\bar{u}_{0,n}\bar{z}_n' \\ &= n^{-(1+\alpha)/2}U_0'\tilde{Z} - \left(\frac{1}{n^{1/2}}\sum_{t=1}^{n} u_{0t}\right)\left(\frac{1}{n^{1+\alpha/2}}\sum_{t=1}^{n} \tilde{z}_{t}\right)' \\ &= n^{-(1+\alpha)/2}U_0'\tilde{Z} + o_p(1), \end{aligned}$$

so both sample moment matrices $n^{-(1+\alpha)}\underline{X}'\tilde{Z}$ and $n^{-(1+\alpha)/2}\underline{U}_0'\tilde{Z}$ are asymptotically equivalent to $n^{-(1+\alpha)}X'\tilde{Z}$ and $n^{-(1+\alpha)}X'\tilde{Z}$ respectively and Theorem 3.7 of PM continues to apply.

Theorem 4.1 follows under the conditions of Lemma 4.1(ii), which guarantee the validity of (4.24).

Proof of Theorem 4.2. The proof will follow by first showing that the "undemeaned" statistic

$$W_n^{(1)} = \left(H\mathrm{vec}\tilde{A}_n - h\right)'\left[H\left\{\left(\underline{X}'P_{\tilde{Z}}\underline{X}\right)^{-1}\otimes\hat{\Omega}_{00}\right\}H'\right]^{-1}\left(H\mathrm{vec}\tilde{A}_n - h\right)$$

has a $\chi^2(r)$ limit distribution under H_0, and then that $W_n^{(1)}$ and W_n are asymptotically equivalent in the sense that

$$W_n - W_n^{(1)} = O_p\left(\frac{1}{n^{1-\beta}}\right). \tag{4.53}$$

We start with the case $\beta < \alpha$. By PM, $n^{-1-\beta}\tilde{Z}'\tilde{Z} \to_p V_{zz}$, so (4.50), (4.51) and (4.52) yield

$$\left(\frac{1}{n^{1+\beta}}\underline{X}'P_{\tilde{Z}}\underline{X}\right)^{-1} = \left(\frac{1}{n^{1+\beta}}\tilde{Z}'\underline{X}\right)^{-1}\left(\frac{1}{n^{1+\beta}}\tilde{Z}'\tilde{Z}\right)\left(\frac{1}{n^{1+\beta}}\underline{X}'\tilde{Z}\right)^{-1} \Rightarrow \Xi_{xx}, \tag{4.54}$$

where $\Xi_{xx} := \left(\tilde{\Psi}_{xx}'\right)^{-1} C_z V_{zz} C_z \tilde{\Psi}_{xx}^{-1}$ and $\tilde{\Psi}_{xx}$ is the random matrix defined in Theorem 4.1. With this notation, the conclusion of Theorem 4.1 (i) becomes

$$n^{\frac{1+\beta}{2}}\text{vec}\left(\tilde{A}_n - A\right) \Rightarrow MN\left(0, \Xi_{xx} \otimes \Omega_{00}\right). \tag{4.55}$$

The Wald statistic in (4.27) can be written as a simple quadratic form: $W_n = \xi_n'\xi_n$, where

$$\xi_n = \left[H\left\{\left(\underline{X}'P_{\tilde{Z}}\underline{X}\right)^{-1} \otimes \hat{\Omega}_{00}\right\}H'\right]^{-1/2}\left(H\text{vec}\tilde{A}_n - h\right).$$

Under the null hypothesis (4.26),

$$\begin{aligned}
\xi_n &= \left[H\left\{\left(\frac{1}{n^{1+\beta}}\underline{X}'P_{\tilde{Z}}\underline{X}\right)^{-1} \otimes \hat{\Omega}_{00}\right\}H'\right]^{-1/2} H\text{vec}n^{-\frac{1+\beta}{2}}\left(\tilde{A}_n - A\right) \\
&\Rightarrow \left[H\left(\Xi_{xx}^{-1} \otimes \hat{\Omega}_{00}\right)H'\right]^{-1/2} MN\left(0, H\left(\Xi_{xx} \otimes \Omega_{00}\right)H'\right) \\
&= N\left(0, I_r\right)
\end{aligned}$$

by (4.54) and (4.55), where the random covariance matrix algebra is justified by mixed normality. This shows Theorem 4.2 for $\beta < \alpha$.

The proof of Theorem 4.2 for $\beta \geq \alpha$ follows an identical argument as is contained in PM since the presence of an intercept in the model does not affect IVX limit theory when $\alpha \in (0,1)$. We have established that, under the assumptions of Theorem 4.2, a $W_n^{(1)} \Rightarrow \chi^2(r)$ under H_0.

It remains to show (4.53). We need to compare $\underline{X}'P_{\tilde{Z}}\underline{X}$ and $\underline{X}'P_{\underline{\tilde{Z}}}\underline{X}$: the identity

$$
\begin{aligned}
\underline{X}'\underline{\tilde{Z}} &= \underline{X}'\tilde{Z} - n\bar{x}_{n-1}\bar{z}'_{n-1} + n\bar{x}_{n-1}\bar{z}'_{n-1} \\
&= \underline{X}'\tilde{Z}
\end{aligned}
$$

yields

$$
\begin{aligned}
\left(\underline{X}'P_{\underline{\tilde{Z}}}\underline{X}\right)^{-1} &= \left[\underline{X}'\tilde{Z}\left(\underline{\tilde{Z}}'\underline{\tilde{Z}}\right)^{-1}\tilde{Z}'\underline{X}\right]^{-1} \\
&= \left(\tilde{Z}'\underline{X}\right)^{-1}\underline{\tilde{Z}}'\underline{\tilde{Z}}\left(\underline{X}'\tilde{Z}\right)^{-1} \\
&= \left(\tilde{Z}'\underline{X}\right)^{-1}\left(\tilde{Z}'\tilde{Z} - n\bar{z}_{n-1}\bar{z}'_{n-1}\right)\left(\underline{X}'\tilde{Z}\right)^{-1} \\
&= \left(\underline{X}'P_{\tilde{Z}}\underline{X}\right)^{-1} - n\left(\tilde{Z}'\underline{X}\right)^{-1}\bar{z}_{n-1}\bar{z}'_{n-1}\left(\underline{X}'\tilde{Z}\right)^{-1}. \qquad (4.56)
\end{aligned}
$$

By Lemma 7.4 (i),

$$
\underline{X}'\tilde{Z} = O_p\left(n^{1+\beta}\right) \quad \bar{z}_{n-1} = O_p\left(n^{\beta-1/2}\right)
$$

so (4.56) implies that

$$\begin{aligned}
\left(\frac{\underline{X}'P_{\underline{\tilde{Z}}}\underline{X}}{n^{1+\beta}}\right)^{-1} - \left(\frac{\underline{X}'P_{\tilde{Z}}\underline{X}}{n^{1+\beta}}\right)^{-1} &= -n^{2+\beta}\left(\tilde{Z}'\underline{X}\right)^{-1}\bar{z}_{n-1}\bar{z}_{n-1}'\left(\underline{X}'\tilde{Z}\right)^{-1} \\
&= O_p\left(n^{2+\beta}n^{-1-\beta}n^{2\beta-1}n^{-1-\beta}\right) \\
&= O_p\left(\frac{1}{n^{1-\beta}}\right),
\end{aligned}$$

showing (4.53) and the theorem.

4.8. Tables and figures

Table 4.1. Unit root tests for the regressors. ADF is the augmented Dickey Fuller test, DF-GLS refers to the Elliot *et al.* (1996) DF-GLS test statistic and PP is the Phillips-Perron statistic. For the ADF and DF-GLS statistics the Bayesian Information Criterion is used.

	ADF	$DF-GLS$	PP
d/e	-3.388^{***}	-3.361^{***}	-2.798^{*}
lty	-1.266	-0.988	-1.341
d/y	-1.966	-1.285	-1.837
d/p	-1.962	-1.304	-1.916
tbl	-2.301	-2.242^{**}	-2.229
e/p	-2.786^{*}	-2.025^{**}	-2.917^{**}
b/m	-3.072^{**}	-2.667^{***}	-2.917^{**}
csp	-2.816^{*}	-1.410	-2.261
dfy	-3.369^{**}	-3.312^{***}	-3.399^{**}
$ntis$	-3.897^{***}	-0.798	-4.293^{***}
tms	-5.170^{***}	-3.901^{***}	-4.709^{***}

*,** and *** imply rejection of the H_0 at 10%, 5% and 1% levels respectively

Table 4.2. Univariate regressions of (CRSP) value weighted returns. t_{NW} refers to the t-ratio statistic with Newey-West HAC standard errors. δ is the long-run correlation coefficient of the residuals from regression models (4.2) and (4.3).

	$\tilde{A}_{OLS}$	t_{NW}	$\tilde{A}_{IVX}$	$Wald$	δ
d/e	-0.00087	-0.0742	-0.00063	0.00777	-0.17604
lty	-0.07592	-1.2635	-0.05620	0.52806	-0.18181
d/y	0.00952	1.6500*	0.01063	5.68342**	-0.83763
d/p	0.00834	1.5440	0.01016	5.22706**	-0.90559
tbl	-0.10191	-1.7369*	-0.08386	1.54609	0.02900
e/p	0.01197	2.7502***	0.01199	5.41182**	-0.66132
b/m	0.01939	1.6963*	0.01993	6.97862***	-0.87121
csp	2.12837	2.8403***	0.34676	0.14463	0.22458
dfy	0.47542	0.7053	0.31200	1.26074	-0.63393
$ntis$	-0.21689	-2.5459**	-0.38653	13.55450***	0.25155
tms	0.20935	1.5497	0.21426	2.00926	-0.14941

* implies rejection at 10% level

** implies rejection at 5% level

*** implies rejection at 1% level

Table 4.3. Univariate regressions with dependent variable being each of the portfolios ordered according to size; first row is $\tilde{A}_{IVX}$, the second is the Wald statistic for individual significance of the regressor, and the third is δ.

Size 1	Size 2	Size 3	Size 4	Size 5	Size 6	Size 7	Size 8	Size 9	Size 10
				d/e					
0.02192	0.01148	0.01046	0.00845	0.00590	0.00737	0.00418	0.00515	−0.00007	−0.00105
2.60996	0.94153	0.97169	0.73580	0.40571	0.67025	0.23909	0.42913	0.00010	0.02486
−0.15550	−0.14992	−0.15706	−0.14494	−0.15409	−0.15419	−0.16513	−0.17923	−0.17272	−0.18487
				lty					
−0.16823	−0.09375	−0.07193	−0.04938	−0.04434	−0.06153	−0.06319	−0.05248	−0.06538	−0.05708
1.30438	0.53297	0.38940	0.21318	0.19446	0.39782	0.46477	0.37882	0.63385	0.62404
−0.07903	−0.11438	−0.11554	−0.13384	−0.14896	−0.16541	−0.15879	−0.18705	−0.20138	−0.19251
				d/y					
0.02448	0.01963	0.01961	0.01893	0.01658	0.01626	0.01387	0.01294	0.01110	0.00894
8.29663***	7.05076***	8.74360***	9.52692***	8.23470***	8.38077***	6.73380***	6.92367***	5.47594**	4.60806**
−0.68725	−0.73527	−0.76529	−0.75965	−0.77622	−0.78128	−0.80047	−0.81255	−0.82360	−0.82695
				d/p					
0.02296	0.01833	0.01850	0.01803	0.01575	0.01554	0.01324	0.01258	0.01063	0.00845
7.45464***	6.26555**	7.93809***	8.77334***	7.54446***	7.74846***	6.22284**	6.60794**	5.07599**	4.13935**
−0.71846	−0.77481	−0.81347	−0.81073	−0.83096	−0.84018	−0.86119	−0.88356	−0.89212	−0.89003
				tbl					
−0.27233	−0.18663	−0.15980	−0.13397	−0.12185	−0.13227	−0.12119	−0.10350	−0.10202	−0.08112
4.59922**	2.81787*	2.56029	2.08829	1.95114	2.44794	2.26769	1.95207	2.03772	1.65651
0.03424	0.02360	0.03906	0.02096	0.01585	0.01252	0.02441	0.00182	0.01105	0.03705
				e/p					
0.01535	0.01523	0.01596	0.01646	0.01508	0.01419	0.01308	0.01190	0.01223	0.01022
2.46483	3.21486*	4.38012**	5.41691**	5.14243**	4.78894**	4.52037**	4.39663**	4.99701**	4.51330**
−0.50884	−0.56330	−0.59070	−0.59585	−0.60891	−0.61656	−0.62922	−0.63931	−0.65210	−0.64233
				b/m					
0.06053	0.04993	0.04600	0.04231	0.03680	0.03512	0.03074	0.02904	0.02364	0.01556
18.95445***	16.94878***	17.84389***	17.48163***	14.79366***	14.05294***	11.90991***	12.47281***	8.77916***	4.88333**
−0.70230	−0.76963	−0.80731	−0.80001	−0.83074	−0.83319	−0.83112	−0.84411	−0.86410	−0.85402
				csp					
0.66040	0.47721	0.12730	0.35916	0.23298	0.12432	0.29386	0.17667	0.22871	0.41061
0.15581	0.11002	0.01021	0.08525	0.04000	0.01244	0.07184	0.03130	0.05930	0.21504
0.23143	0.23849	0.24796	0.25883	0.27735	0.26294	0.26392	0.24848	0.23239	0.22316
				dfy					
1.80304	1.42976	1.28304	1.14677	0.96380	0.89708	0.74602	0.70621	0.46958	0.18845
11.65671***	9.63105***	9.58927***	8.88326***	7.10993***	6.48543**	5.02004**	5.29776**	2.51989	0.53280
−0.62381	−0.64973	−0.66314	−0.65583	−0.66553	−0.66211	−0.63352	−0.63969	−0.64385	−0.60093
				$ntis$					
−0.91804	−0.80956	−0.72245	−0.66123	−0.60441	−0.58206	−0.55156	−0.49601	−0.47532	−0.34440
21.31776***	22.82548***	22.72286***	21.82350***	20.90155***	19.97675***	20.46489***	19.42084***	18.99639***	12.26169***
0.14040	0.16391	0.15470	0.18306	0.18936	0.20006	0.20946	0.19742	0.25215	0.24129
				tms					
0.74857	0.59338	0.53964	0.49341	0.44959	0.43458	0.37471	0.32580	0.26848	0.19676
6.89480***	5.75309**	5.91830**	5.75079**	5.42706**	5.33228**	4.38560**	3.90334*	2.82205*	1.93790
−0.07864	−0.08782	−0.10704	−0.09733	−0.09857	−0.10766	−0.11816	−0.11062	−0.13488	−0.16689

Table 4.4. Univariate regressions with dependent variable being each of the portfolios ordered according to book-to-market value; first row is $\tilde{A}_{IVX}$, the second is the Wald statistic for the significance of the regressor and the third is δ.

Value 1	Value 2	Value 3	Value 4	Value 5	Value 6	Value 7	Value 8	Value 9	Value 10
				d/e					
0.00033	0.00161	0.00030	−0.00058	−0.00397	−0.00441	0.00021	0.00228	0.00222	0.00959
0.00185	0.05847	0.00196	0.00527	0.27010	0.28894	0.00057	0.06428	0.05272	0.67174
−0.14848	−0.14501	−0.14371	−0.17526	−0.15030	−0.18611	−0.13750	−0.15572	−0.14823	−0.15562
				lty					
−0.09589	−0.04528	−0.04481	−0.04282	−0.06890	−0.04509	−0.04617	−0.05425	−0.04516	−0.03434
1.32829	0.39544	0.38234	0.24291	0.69304	0.25563	0.24192	0.31250	0.18643	0.07273
−0.15996	−0.21734	−0.24136	−0.17509	−0.18429	−0.17744	−0.17522	−0.16191	−0.14040	−0.11436
				d/y					
0.00940	0.01013	0.00910	0.01113	0.00957	0.01040	0.01259	0.01489	0.01692	0.02237
3.80456*	5.95034**	4.75407**	4.96483**	4.01458**	4.09323**	5.43914**	7.09458***	7.94526***	9.44250***
−0.79045	−0.82013	−0.82198	−0.80768	−0.77583	−0.74489	−0.76658	−0.74855	−0.75077	−0.73627
				d/p					
0.00883	0.00979	0.00872	0.01049	0.00901	0.00973	0.01214	0.01439	0.01619	0.02148
3.39054*	5.60671**	4.40740**	4.46989**	3.59004*	3.63674*	5.10817**	6.70272***	7.35103***	8.83509***
−0.84882	−0.88852	−0.88639	−0.86730	−0.83499	−0.80752	−0.83422	−0.81507	−0.80525	−0.78891
				tbl					
−0.10958	−0.07053	−0.07698	−0.08807	−0.09149	−0.09216	−0.09119	−0.10418	−0.10812	−0.13322
2.27072	1.25957	1.48581	1.35147	1.60879	1.40959	1.24294	1.52360	1.41088	1.44584
0.08189	−0.01819	−0.03013	0.00626	0.00430	0.01392	−0.02529	−0.00779	0.00897	0.02698
				e/p					
0.00995	0.01046	0.00982	0.01228	0.01226	0.01335	0.01381	0.01540	0.01748	0.01982
3.20959*	4.75881**	4.16102**	4.55173**	4.96367**	5.10167**	4.91484**	5.70586**	6.37554**	5.58956**
−0.63075	−0.66623	−0.66652	−0.62946	−0.61838	−0.57161	−0.62577	−0.59525	−0.59195	−0.57080
				b/m					
0.01629	0.01794	0.01674	0.02588	0.02221	0.02558	0.02952	0.03661	0.03641	0.04999
4.03439**	6.57321**	5.68142**	9.46745***	7.58467***	8.74489***	10.52765***	15.43051***	13.15414***	17.40796***
−0.79889	−0.82454	−0.82156	−0.90433	−0.86630	−0.88229	−0.83625	−0.81629	−0.79093	−0.75403
				csp					
0.55496	0.27176	0.11150	0.17779	0.28843	0.10094	0.23841	0.43759	0.59565	−0.04803
0.27810	0.08568	0.01422	0.03633	0.10592	0.01233	0.06033	0.18832	0.28258	0.00113
0.28535	0.26157	0.20776	0.20585	0.14909	0.17817	0.20118	0.14920	0.15076	0.17147
				dfy					
0.18446	0.36367	0.33211	0.60348	0.27685	0.41725	0.75745	0.86777	0.89209	1.41746
0.38386	1.99473	1.65021	3.75251*	0.86579	1.67796	4.96162**	6.03304**	5.56101**	9.53358***
−0.54529	−0.55006	−0.55952	−0.65114	−0.64841	−0.68364	−0.68749	−0.68405	−0.67214	−0.68942
				$ntis$					
−0.38359	−0.35270	−0.33352	−0.49026	−0.46532	−0.49181	−0.57492	−0.58413	−0.62136	−0.70994
11.31409***	12.93359***	11.66306***	17.93826***	17.37163***	16.72782***	21.63509***	20.42904***	20.06409***	18.04026***
0.18933	0.20914	0.19151	0.30029	0.35292	0.33296	0.31650	0.24058	0.21885	0.13560
				tms					
0.20312	0.19306	0.22874	0.28878	0.19765	0.28888	0.28034	0.31456	0.35742	0.53443
1.52736	1.87925	2.64755	2.95165*	1.48821	2.78299*	2.37210	2.76782*	3.09734*	4.75900**
−0.21012	−0.12091	−0.11603	−0.11085	−0.11670	−0.11861	−0.06678	−0.08287	−0.08202	−0.08601

Table 4.5. Multivariate regressions of (CRSP) value weighted returns.

$\tilde{A}_{IVX}$								$Wald$
d/e	d/y	tbl	e/p	b/m	dfy	$ntis$	tms	
−0.08874	0.07721	−0.19299	−0.07120	0.01188	0.23468	−0.19933	0.02313	29.10761***
	0.00299		−0.00002	0.01332		−0.28835		16.77229***
−0.02065		−0.11888			0.67978		0.11675	7.07384
	0.00496		0.00786					6.33305**
	0.00011			0.01977				7.03630**
	0.00991					−0.30931		15.68920***
			0.00394	0.01603				7.67070**
			0.00874			−0.35595		16.58195***
				0.01780		−0.28265		14.85310***
	−0.00380		0.00475	0.02071				7.78093*
	0.00875		0.00171			−0.31164		16.83297***
	0.00276			0.01355		−0.28924		15.52520***
			0.00089	0.01692		−0.28590		16.42462***
−0.00623		−0.10579						2.01718
−0.00897					0.49703			2.26715
−0.00572							0.26937	2.97177
		−0.07983			0.31685			2.82219
		−0.04658					0.16444	2.44297
					0.21889		0.16351	2.32956
−0.02111		−0.14865			0.75182			6.76303*
−0.00832		−0.06117					0.22521	3.75598
−0.01214					0.43270		0.22627	4.40028
		−0.06059			0.26484		0.08819	3.02176

Table 4.6. Multivariate estimation of the returns of portfolios ordered according to size. The Wald statistic tests the hypothesis that each regressor is insignificant for all the portfolios jointly.

				$\tilde{A}'_{IVX}$						$Wald$
Size 1	Size 2	Size 3	Size 4	Size 5	Size 6	Size 7	Size 8	Size 9	Size 10	
				d/e						
0.02192	0.01148	0.01046	0.00845	0.00590	0.00737	0.00418	0.00515	−0.00007	−0.00105	28.16805***
				lty						
−0.16823	−0.09375	−0.07193	−0.04938	−0.04434	−0.06153	−0.06319	−0.05248	−0.06538	−0.05708	12.05108
				d/y						
0.02448	0.01963	0.01961	0.01893	0.01658	0.01626	0.01387	0.01294	0.01110	0.00894	22.43189**
				d/p						
0.02296	0.01833	0.01850	0.01803	0.01575	0.01554	0.01324	0.01258	0.01063	0.00845	22.18228**
				tbl						
−0.27233	−0.18663	−0.15980	−0.13397	−0.12185	−0.13227	−0.12119	−0.10350	−0.10202	−0.08112	12.76743
				e/p						
0.01535	0.01523	0.01596	0.01646	0.01508	0.01419	0.01308	0.01190	0.01223	0.01022	15.56151
				b/m						
0.06053	0.04993	0.04600	0.04231	0.03680	0.03512	0.03074	0.02904	0.02364	0.01556	32.84576***
				csp						
0.66040	0.47721	0.12730	0.35916	0.23298	0.12432	0.29386	0.17667	0.22871	0.41061	3.71252
				dfy						
1.80304	1.42976	1.28304	1.14677	0.96380	0.89708	0.74602	0.70621	0.46958	0.18845	29.34962***
				$ntis$						
−0.91804	−0.80956	−0.72245	−0.66123	−0.60441	−0.58206	−0.55156	−0.49601	−0.47532	−0.34440	25.53436***
				tms						
0.74857	0.59338	0.53964	0.49341	0.44959	0.43458	0.37471	0.32580	0.26848	0.19676	10.41352

Table 4.7. Multivariate estimation of the returns of portfolios ordered according to the book-to-market value. The Wald statistic tests the hypothesis that each regressor is insignificant for all the portfolios jointly.

				$\tilde{A}'_{IVX}$						$Wald$
Value 1	Value 2	Value 3	Value 4	Value 5	Value 6	Value 7	Value 8	Value 9	Value 10	
				d/e						
0.00033	0.00161	0.00030	−0.00058	−0.00397	−0.00441	0.00021	0.00228	0.00222	0.00959	10.04680
				lty						
−0.09589	−0.04528	−0.04481	−0.04282	−0.06890	−0.04509	−0.04617	−0.05425	−0.04516	−0.03434	5.13271
				d/y						
0.00940	0.01013	0.00910	0.01113	0.00957	0.01040	0.01259	0.01489	0.01692	0.02237	14.65000
				d/p						
0.00883	0.00979	0.00872	0.01049	0.00901	0.00973	0.01214	0.01439	0.01619	0.02148	14.47177
				tbl						
−0.10958	−0.07053	−0.07698	−0.08807	−0.09149	−0.09216	−0.09119	−0.10418	−0.10812	−0.13322	3.40673
				e/p						
0.00995	0.01046	0.00982	0.01228	0.01226	0.01335	0.01381	0.01540	0.01748	0.01982	7.92923
				b/m						
0.01629	0.01794	0.01674	0.02588	0.02221	0.02558	0.02952	0.03661	0.03641	0.04999	28.06062***
				csp						
0.55496	0.27176	0.11150	0.17779	0.28843	0.10094	0.23841	0.43759	0.59565	−0.04803	4.86997
				dfy						
0.18446	0.36367	0.33211	0.60348	0.27685	0.41725	0.75745	0.86777	0.89209	1.41746	36.98133***
				$ntis$						
−0.38359	−0.35270	−0.33352	−0.49026	−0.46532	−0.49181	−0.57492	−0.58413	−0.62136	−0.70994	25.15749***
				tms						
0.20312	0.19306	0.22874	0.28878	0.19765	0.28888	0.28034	0.31456	0.35742	0.53443	10.30016

Table 4.8. Subsample estimation of univariate regressions. CRSP is used as the dependent variable. The Wald and the t_{NW} statistics test the individual significance of the regressor in each case.

$\tilde{A}_{OLS}$		t_{NW}		$\tilde{A}_{IVX}$		$Wald$		δ	
Panel A	Panel B	Panel A	Panel B	Panel A	Panel B	Panel A	Panel B	Panel A	Panel B
				d/e					
−0.01275	−0.00014	−0.5392	−0.0171	−0.01232	0.00145	0.89000	0.02119	−0.22329	−0.04343
				lty					
−0.44115	−0.01884	−1.4537	−0.2208	−0.20587	−0.03006	0.21390	0.09374	−0.00246	−0.33097
				d/y					
0.02189	0.00428	1.2563	0.9196	0.01988	0.00342	3.53511*	0.51508	−0.82939	−0.85518
				d/p					
0.01674	0.00421	1.0831	0.8966	0.01690	0.00338	2.64318	0.50063	−0.90348	−0.96406
				tbl					
−0.18582	−0.09041	−0.9475	−1.4290	−0.08208	−0.08756	0.12431	1.33756	0.13019	−0.09158
				e/p					
0.02299	0.00463	2.6649***	0.8918	0.02189	0.00298	4.73702**	0.35100	−0.72646	−0.63640
				b/m					
0.04942	0.00055	1.8368*	0.0733	0.04451	−0.00075	8.47218***	0.01043	−0.90469	−0.81395
				csp					
2.26073	2.55827	2.1355**	1.8218*	0.62654	0.53382	0.29322	0.05498	0.14505	0.40407
				dfy					
0.36090	0.90456	0.4383	1.6078	0.22197	0.34213	0.37341	0.44752	−0.75591	−0.23833
				$ntis$					
−0.28515	−0.19968	−1.9850**	−2.1958**	−0.39349	−0.17917	8.08120***	2.69014	0.26850	0.07259
				tms					
0.16396	0.26070	0.4373	2.1292**	0.06208	0.25303	0.03207	3.81326*	−0.15701	−0.11067

Panel A: January 1927 - June 1967. For *csp* July 1937 - June 1967

Panel B: July 1967 - December 2008. For *csp* July 1937 - December 2002

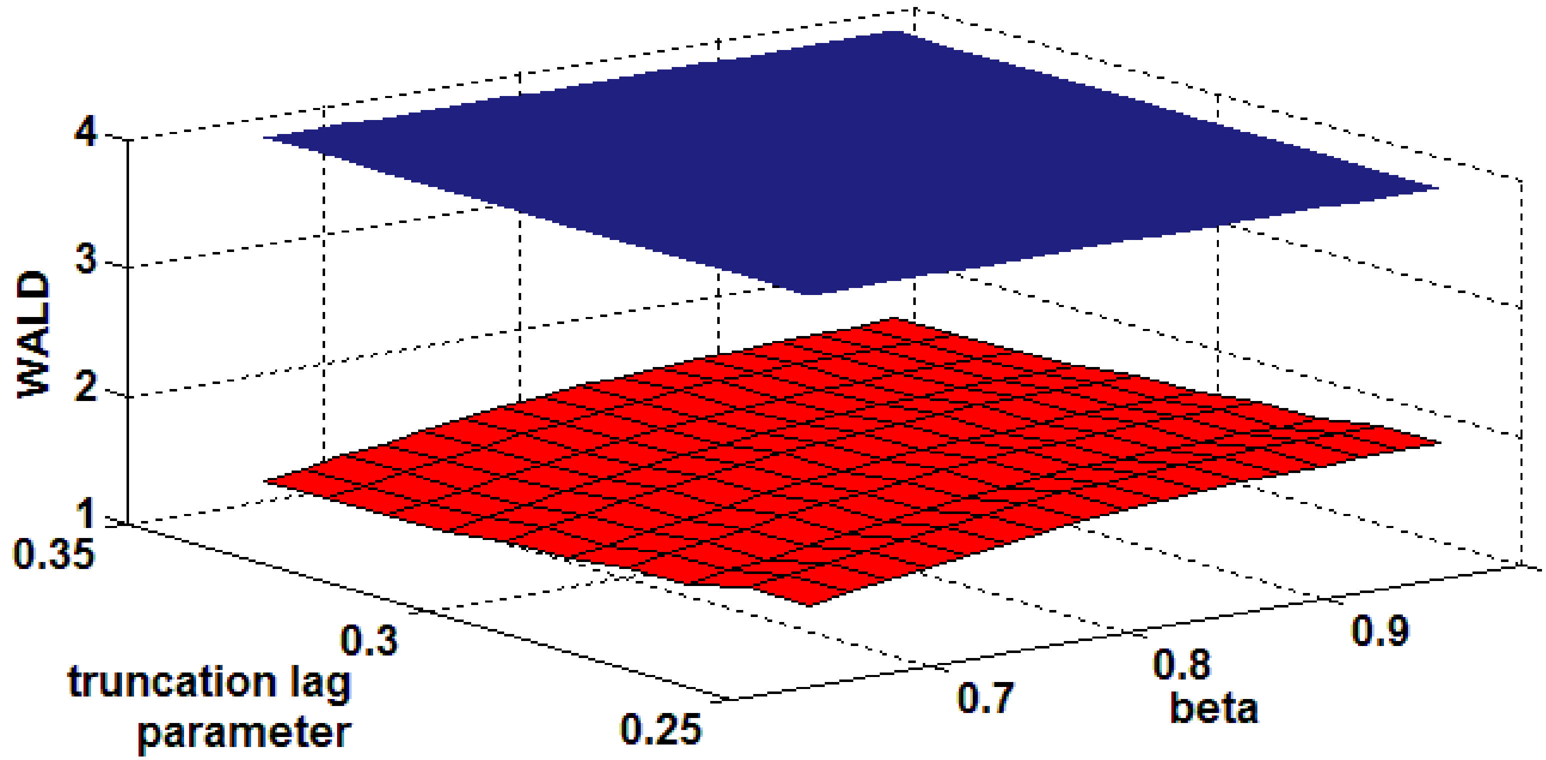
Table 4.9. Subsample estimation of multivariate regressions. CRSP is the dependent variable; first row of each case refers to Panel A and second row refers to Panel B; Wald statistic tests the overall significance of the regressors.

$\tilde{A}_{IVX}$								$Wald$
d/e	d/y	tbl	e/p	b/m	dfy	$ntis$	tms	
−0.21665	0.11845	−0.89969	−0.16769	0.10762	1.25304	0.03582	−1.67522	39.19989***
0.05953	−0.04139	−0.47307	0.09718	−0.06681	2.24342	−0.01519	−0.19650	24.01531***
−0.21329	0.13064	0.02875	−0.16090	0.09677	0.74755	0.02208		37.06066***
0.04963	−0.03881	−0.37402	0.08767	−0.05651	1.76481	−0.02757		22.63974***
−0.21198	0.12879		−0.15923	0.09559	0.74841	0.02793		35.81690***
0.03335	−0.02170		0.04666	−0.03960	0.92807	−0.05160		7.07848
	−0.04794		0.02975	0.07340	0.16280	−0.15207		20.52097***
	0.01090		0.01297	−0.03775	0.88233	−0.05101		7.32062
	−0.04280		0.02514	0.07431		−0.16220		19.95571***
	0.01258		0.00380	−0.01958		−0.15299		5.11528
−0.06581		−0.04035			1.76876		−0.08303	9.50316**
−0.01166		−0.09299			0.79795		0.23011	7.72917
		0.10951				−0.42942		10.74334***
		−0.10151				−0.21259		4.93665*

Panel A: January 1927 - June 1967, Panel B: July 1967 - December 2008

Figure 4.1. H_0 : insignificance of tbl, sensitivity of Wald statistic to β and γ.
sensitivity tbl

(2)



$_1 2.pdf$

Figure 4.2. H_0 : insignificance of dy, sensitivity of Wald statistic to β and γ.

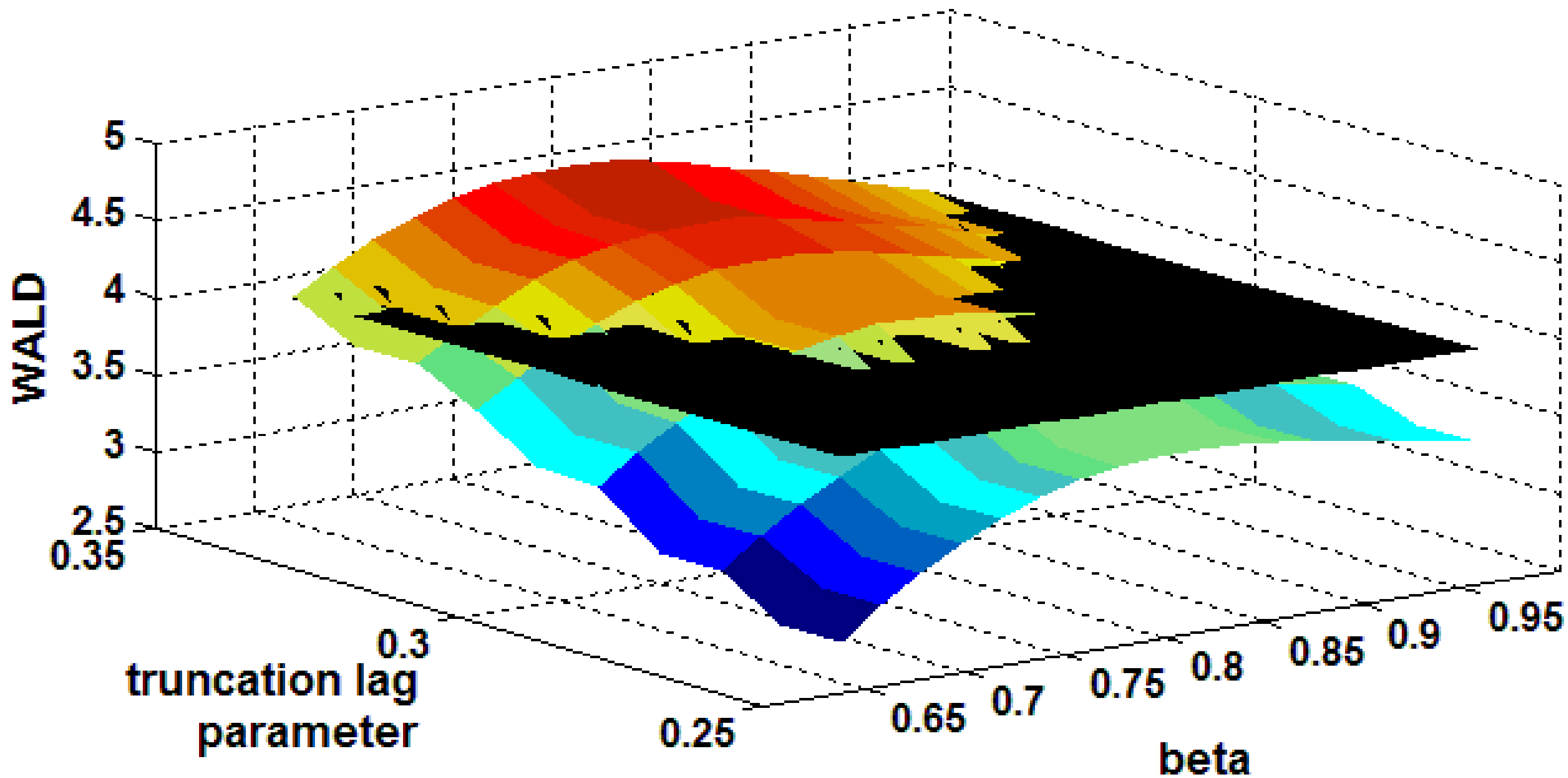
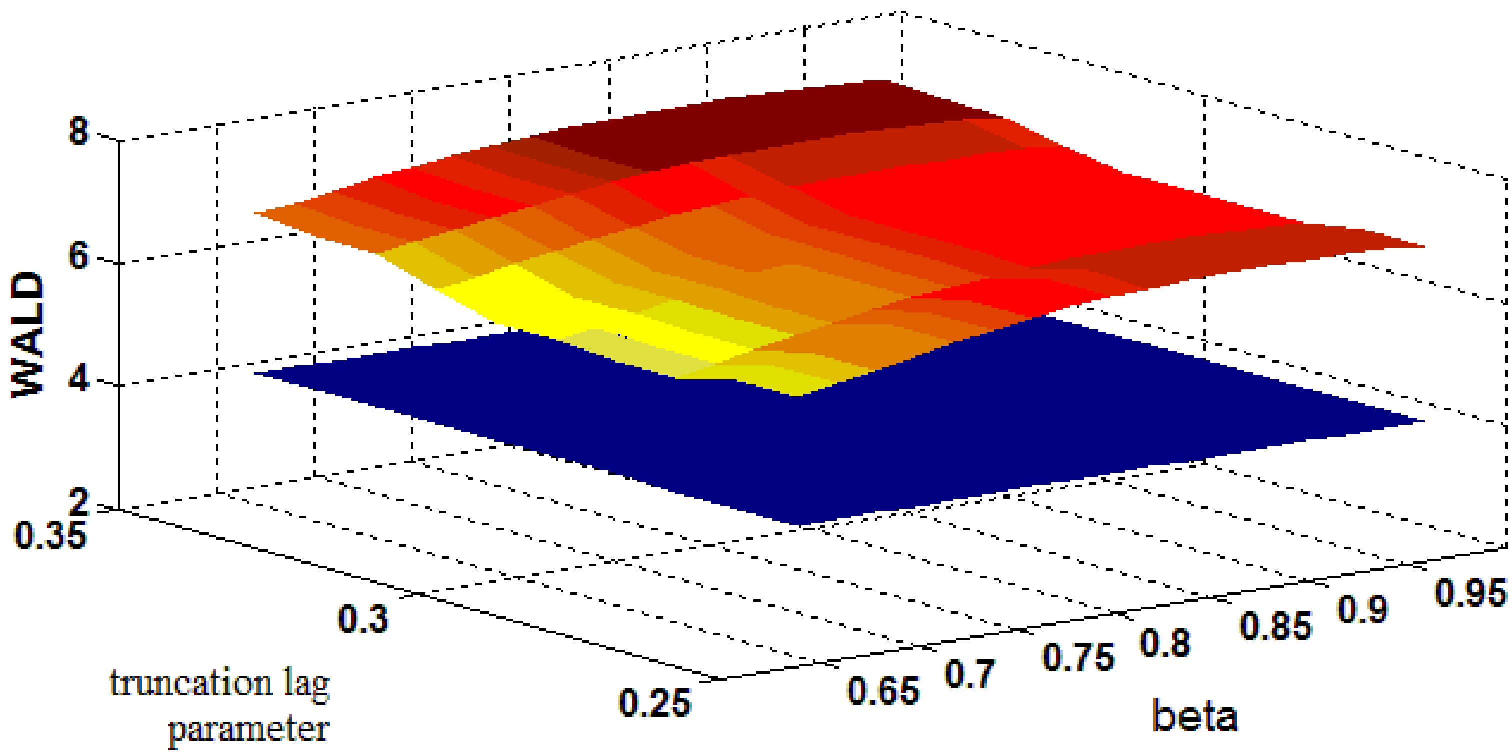


Figure 4.3. H_0 : insignificance of b/m, sensitivity of Wald statistic to β and γ.



CHAPTER 5

Conclusion

The present thesis discusses econometric inference under a variety of nonstationary frameworks.

In Chapter 2 we perform an extensive simulation study for the finite sample properties of the OLS estimator in vector autoregressive models. We broaden the scope of the results by Abadir, Hadri and Tzavalis (1999) by introducing overparameterized models including deterministic components and excessive lag terms. Their scalar bias matrix result is generalised to the overparameterized case. We apply the response surface methodology to derive numerical approximations for the bias and variance. In the absence of analytical results for finite samples these approximations can be valuable for practitioners using vector autoregressions.

In Chapter 3 we generalise the BNM and BEPO test statistics proposed by Forchini and Marsh (2000) to account for autocorrelation in the error term. Autocorrelation is introduced in the form of a finite order moving average process and is accounted for in the construction of the test statistics. Therefore, the resulting BNM and BEPO tests are free of nuisance parameters. The feasibility of our procedure is achieved by maximum likelihood estimation of the moving average parameters and by the use of information criteria for moving average lag order determination. Comparing the finite sample properties of our generalised BNM and BEPO tests with the ones originally proposed by Forchini and Marsh, we achieve an enormous reduction of size distortion in the presence of autocorrelated errors. In the absence of autocorrelated errors our simulation study suggests that the power loss of the generalised statistics is relatively small. We also compare the generalised BNM and BEPO statistics to the statistics derived by Ng and Perron (2001), Perron and Qu (2007) and Seo (2006) and observe that our statistics exhibit small size and high power. Our simulation experiments reveal serious problems associated with the finite sample properties of the Ng and Perron (2001) statistics: power reversal, power non-monotonicity with respect to the sample size for some alternatives and extremely low power (lower than the nominal size) in some cases. A further observation is that the power reversal problem is not eliminated neither by Seo (2006), who first reported it, nor from Perron and Qu (2007) who attempted to solve it. What makes our tests stand out is their robustness to the presence of autocorrelation in the errors and their improved performance: their size is comparatively low and reduces substantially as sample size increases and they have high power across a variety of alternatives which always increases with the sample size.

Chapter 4 makes a methodological contribution to testing the hypothesis of predictability of stock returns. The IVX methodology of Phillips and Magdalinos (2009) is modified and extended in order to apply to a system of predictive regressions with an intercept. The proposed approach has two main advantages over existing methods. First, it provides inference that does not depend on *a priori* knowledge of the degree of regressor persistence. Second, it easily accommodates joint inference in multiple predictive regression models. The importance of assessing the combined predictive power of a set of explanatory variables is an interesting empirical finding. In one characteristic example (testing predictability for the second half of our sample), all available explanatory variables appear to be individually insignificant (at the 5% level) as predictors of the market portfolio as a result of performing a individual hypothesis test for each explanatory variable. However, a joint hypothesis test for the same variables leads to strong rejection of the null hypothesis of no predictability (even at the 1% level). In this example, while each explanatory variable has limited predictive value, their combination has significant predictive power. The methodology of Chapter 4 addresses important issues for applied research in predictive regressions by extending both the validity of inference (by accommodating a large class of persistent regressors) and the range of testable hypotheses to include general linear restrictions on a multivariate regression framework.

References

[1] Abadir, K. M. (1993), OLS bias in a nonstationary autoregression, *Econometric Theory* 9, 81-93.

[2] Abadir, K. M. (1995), Unbiased estimation as a solution to testing for random walks, *Economics Letters* 47, 263-268.

[3] Abadir, K. M. and K. Hadri (2000), Is more information a good thing? Bias nonmonotonicity in stochastic difference equations, *Bulletin of Economic Research* 52, 91-100.

[4] Abadir, K. M., Hadri, K. and E. Tzavalis (1999), The influence of VAR dimensions on estimator biases, *Econometrica* 67, 163-181.

[5] Abadir, K. M. and R. Larsson (1996), The joint moment generating function of quadratic forms in multivariate autoregressive series, *Econometric Theory* 12, 682-704.

[6] Abadir, K. M. and R. Larsson (2001), The joint moment generating function of quadratic forms in multivariate autoregressive series: The case with deterministic components, *Econometric Theory* 17, 222-246.

[7] Abadir, K.M. and J. R. Magnus (2005), *Matrix algebra, volume 1 of econometric exercises*, Cambridge University Press, Cambridge.

[8] Abadir, K. M. and P. Paruolo (2009), On efficient simulations in dynamic models, Chapter 11 (268-299) in Castle J. L. and N. Shephard (eds) *The Methodology and Practice of Econometrics - A Festschrift in Honour of David F. Hendry*, Oxford University Press

[9] Akaike, H. (1974), A new look at the statistical model identification, *I.E.E.E. Transactions on Automatic Control*, AC 19, 716-723.

[10] Anderson, T. W. (1971), *The Statistical Analysis of Time Series*, John Wiley and Sons, New York.

[11] Ang, A. and G. Bekaert (2007), Stock return predictability: Is it there? *Review of Financial Studies*, 20, 651-707.

[12] Avramov, D. (2002), Stock Return Predictability and Model Uncertainty, *Journal of Financial Economics*, 64(3), 423–458.

[13] Banerjee, A., Dolado, J. J., Galbraith, J. W. and D. F. Hendry (1993), *Co-integration, Error-Correction, and the Econometric Analysis of Non-Stationary Data,* Oxford University Press.

[14] Boudoukh, J., Michaely, R., Richardson, M. P. and M. R. Roberts (2007), On the importance of measuring payout yield: implications for empirical asset pricing, *Journal of Finance*, 62(2), 877–915.

[15] Campbell, J. Y. (1987), Stock returns and the Term Structure. *Journal of Financial Economics*, 18, 373–399.

[16] Campbell, J.Y.and R.J. Shiller (1988), Stock prices, earnings, and expected dividends, *Journal of Finance*, 43, 661–676.

[17] Campbell J.Y. and S.B. Thompson (2008), Predicting the Equity Premium Out of Sample: Can anything Beat the Historical Average?, *Review of Financial Studies* 21, 1509-1531.

[18] Campbell J.Y. and M. Yogo (2006), Efficient tests of stock return predictability, *Journal of Financial Economics* 81, 27-60.

[19] Campos, J. (1986), Finite-sample properties of the instrumental-variables estimator for dynamic simultaneous-equation subsystems with ARMA disturbances, *Journal of Econometrics* 32, 333-366.

[20] Cavanagh, C., G. Elliott, and J.H. Stock (1995), Inference in models with nearly integrated regressors, *Econometric Theory* 11, 1131–1147.

[21] Cheung, Y. -W. and K. S. Lai (1995), Lag order and critical values of the augmented Dickey-Fuller test, *Journal of Business and Economic Statistics* 13, 277-280.

[22] Christopherson, J.A., Ferson, W.E., and D.A. Glassman (1998), Conditioning manager alphas on economic information: Another look at the persistence of performance, *Review of Financial Studies* 11, 111-142.

[23] Cochrane, J. (1999), New facts in finance, *Economic Perspectives* 23, 36-58.

[24] Cox, D. R. and D. V. Hinkley (1974), *Theoretical Statistics,* Chapman and Hall, London.

[25] DeJong, D. N., Nankervis, J. C., Savin, N.E. and C. H. Whiteman (1992), The power problems of unit root tests in time series with autocorrelated errors, *Journal of Econometrics* 53, 323-343.

[26] Dickey, D. and W. Fuller (1979), Distribution of the estimators for autoregressive time series with a unit root, *Journal of the American Statistical Association* 74, 427-431.

[27] Dickey, D. and W. Fuller (1981), Likelihood ratio statistics for autoregressive time series with a unit root, *Econometrica* 49, 1057-1072.

[28] Doornik, J. A. and D. F. Hendry (2007), Interactive Monte Carlo experimentation in econometrics using PcNaive4, Timberlake Consultants, London.

[29] Elliott, G. (1998), On the robustness of cointegration methods when regressors almost have unit roots, *Econometrica* 66, 149-158.

[30] Elliott, G. and J.H. Stock (1994), Inference in time series regression when the order of integration of a regressor is unknown, *Econometric Theory*, 10, 672–700.

[31] Elliot, G, Rothenberg, T.J. and J.H. Stock (1996), Efficient tests for an autoregressive unit root, *Econometrica* 64, 813-836.

[32] Engle, R. F. and C. W. J. Granger (1987), Cointegration and error correction: Representation, estimation and testing, *Econometrica* 51, 251-276.

[33] Engle, R. F., Hendry, D. F. and D. Trumble (1985), Small-sample properties of ARCH estimators and tests, *Canadian Journal of Economics* 18, 66-93.

[34] Ericsson, N. R. (1991), Monte Carlo methodology and the finite sample properties of instrumental variables statistics for testing nested and non-nested hypotheses, *Econometrica* 59, 1249-1277.

[35] Ericsson, N. R. and J. G. MacKinnon (2002), Distributions of error correction tests for cointegration, *Econometrics Journal* 5, 285-318.

[36] Evans, G. B. A. and N. E. Savin (1981), Testing for unit roots: 1, *Econometrica* 49, 753-779.

[37] Fama, E.F. (1970), Efficient capital markets: A review of theory and empirical work, *Journal of Finance* 25, 383-417.

[38] Fama, E.F. (1991), Efficient capital markets: II, *Journal of Finance* 46, 1575–1617.

[39] Fama, E.F. and K.R. French (1988), Dividend yields and expected stock returns, *Journal of Financial Economics* 22, 3–24.

[40] Fama, E.F. and K.R. French (1989), Business conditions and expected returns on stocks and bonds. *Journal of Financial Economics* 25, 23–49.

[41] Fama, E.F. and K.R. French (1992), The cross-section of expected stock returns, *Journal of Finance* 47, 427-465.

[42] Fama, E.F. and K.R. French (1993), Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3-56.

[43] Ferson, W.E. and C.R. Harvey (1999), Conditioning variables and the cross section of stock returns, *Journal of Finance* 54, 1325-1360.

[44] Forchini, G. F. and P. Marsh (2000), Exact inference for the unit root hypothesis, Discussion Paper 00/54, University of York .

[45] Fuller, W. A. (1976), *Introduction to Statistical Time Series,* Wiley, New York.

[46] Giraitis, L. and P. C. B. Phillips (2006), Uniform limit theory for stationary autoregression, *Journal of Time Series Analysis* 27, 51-60.

[47] Gonzalo, J. and J.-Y Pitarakis (1998), On the exact moments of asymptotic distributions in an unstable AR(1) with dependent errors, *International Economic Review* 39, 71–88.

[48] Gonzalo, J. and J.-Y Pitarakis (2009), Regime specific predictability in predictive regressions, Discussion Paper, University of Southampton.

[49] Goyal, A. and I Welch (2003), Predicting the equity premium with dividend ratios, *Management Science* 49, 639-654.

[50] Goyal, A. and I. Welch (2008), A Comprehensive Look at the Empirical Performance of Equity Premium Prediction, *Review of Financial Studies* 21, 1455-1508.

[51] Granger, C. W. J. (1981), Some properties of time series data and their use in econometric model specification, *Journal of Econometrics* 16, 121–130.

[52] Granger, C. W. J. and P. Newbold (1974), Spurious regressions in econometrics, *Journal of Econometrics* 2, 111-120.

[53] Hamilton, J. D. (1994), *Time Series Analysis,* Princeton University Press, New Jersey.

[54] Hannan, E. J., and B. G. Quinn (1979), The determination of the order of an autoregression, *Journal of the Royal Statistical Society, Series B* 41, 190-195.

[55] Hendry, D. F. (1984), Monte Carlo experimentation in econometrics, in: Griliches, Z., and M. D. Intriligator (eds) *Handbook of Econometrics*, Volume 2, North-Holland, Amsterdam.

[56] Hendry, D. F. and H.-M. Krolzig (2005), The properties of automatic Gets Modelling, *Economic Journal* 115, C32-C61.

[57] Hillier, G. H. (1987), Classes of similar regions and their power properties for some econometric testing problems, *Econometric Theory* 3, 1-44.

[58] Hodrick, R. J. (1992), Dividend yields and expected stock returns: alternative procedures for inference and measurement. *Review of Financial Studies* 5, 257–86.

[59] Horn R.A. and C.R. Johnson (1985), *Matrix Analysis.* Cambridge University Press, Cambridge

[60] Johansen, S. (1988), Statistical analysis of cointegration vectors, *Journal of Economic Dynamics and Control* 12, 231-254.

[61] Kandel, S. and R.F. Stambaugh (1996), On the predictability of stock returns: An asset allocation perspective, *Journal of Finance* 51, 385-424.

[62] Keim, D.B. and R. F. Stambaugh (1986), Predicting returns in the stock and the bond markets, *Journal of Financial Economics* 17, 357-390.

[63] Kirby, C. (1998), The restrictions on predictability implied by rational asset pricing models, *Review of Financial Studies* 11, 343-382.

[64] Kiviet, J. F. and G. D. A. Phillips (2005), Moment approximation for least-squares estimators in dynamic regression models with a unit root, *Econometrics Journal* 8, 115-142.

[65] Kong, A., Rapach, D.E., Strauss, J.K., Tu, J. and G. Zhou (2009), How predictable are components of the aggregate market portfolio?, Working paper, Available at SSRN: http://ssrn.com/abstract=1307420.

[66] Kothari, S. and J. Shanken (1997), Book-to-market, dividend yield, and expected market returns: a time-series analysis, *Journal of Financial Economics* 44,169–203.

[67] Lamont, O. (1998), Earnings and expected returns. *Journal of Finance* 53,1563–1587.

[68] Lanne, M. (2002), Testing the predictability of stock returns, *Review of Economics and Statistics* 84, 407–415.

[69] Lawford, S. (2001), *Improved modelling in finite-sample and non-linear frameworks,* unpublished.D.Phil. thesis, University of York.

[70] Lawford, S. and M.P. Stamatogiannis (2009), The finite-sample effects of VAR dimensions on OLS bias, OLS variance, and minimum MSE estimators, *Journal of Econometrics* 148, 124–130.

[71] Le Breton, A. and D. T. Pham (1989), On the bias of the least squares estimator for the first order autoregressive process, *Annals of the Institute of Statistical Mathematics* 41, 555-563.

[72] Lehmann, E. L. (1986), *Testing Statistical Hypotheses,* Wiley, New York.

[73] Lehmann, E. L. and C. Stein (1948), Most powerful tests of composite hypotheses, *The Annals of Mathematical Statistics* 19, 495-516.

[74] Lettau, M. and S. Ludvigson (2001), Consumption, aggregate wealth, and expected stock returns, *Journal of Finance* 56, 815-849.

[75] Lettau, M. and S. Ludvigson (2005), Expected returns and expected dividend growth. *Journal of Financial Economics* 76, 583–626.

[76] Lettau, M. and S. van Nieuwerburgh (2008), Reconciling the return predictability evidence, *Review of Financial Studies* 21, 1607-1652.

[77] Lewellen, J. (2004), Predicting returns with financial ratios, *Journal of Financial Economics* 74, 209–235.

[78] Maddala, G. S. and I. M. Kim (1998), *Unit Roots, Cointegration, and Structural Change,* Cambridge University Press, Cambridge.

[79] MacKinnon, J. G. (1994), Approximate asymptotic distribution functions for unit-root and cointegration tests, *Journal of Business and Economic Statistics* 12, 167-176.

[80] MacKinnon, J. G. (1996), Numerical distribution functions for unit root and cointegration tests, *Journal of Applied Econometrics* 11, 601-618.

[81] MacKinnon, J. G., Haug, A. A. and L. Michelis (1999), Numerical distribution functions of likelihood ratio tests for cointegration, *Journal of Applied Econometrics* 14, 563-577.

[82] MacKinnon, J. G. and A. A. Smith, Jr. (1998), Approximate bias correction in econometrics, *Journal of Econometrics* 85, 205-230.

[83] Maeshiro, A. (1999), A lagged dependent variable, autocorrelated disturbances, and unit root tests – peculiar OLS bias properties – a pedagogical note, *Applied Economics* 31, 381-396.

[84] Magdalinos, T. and P. C. B Phillips (2009), Limit theory for cointegrated systems with moderately integrated and moderately explosive regressors, *Econometric Theory* 25, 482-526.

[85] Marsh, P. (2005), A measure of discrimination for the unit root hypothesis, Discussion Paper 05/02, University of York .

[86] Marsh, P. (2007), Constructing optimal tests on a lagged dependent variable, *Journal of Time Series Analysis* 28, 723-743.

[87] Nankervis, J. C. and N. E. Savin (1988), The exact moments of the least squares estimator for the autoregressive model: Corrections and extensions, *Journal of Econometrics* 37, 381-388.

[88] Nelson, C. R. and C. I. Plosser (1982), Trends and random walks in macroeconomic time series: Some evidence and implications, *Journal of Monetary Economics* 10, 139-162.

[89] Ng, S. and P. Perron (2001), Lag length selection and the construction of unit root tests with good size and power, *Econometrica* 69, 1519–1554.

[90] Park, J. Y. and P. C. B. Phillips (1988), Statistical inference in regressions with integrated processes: Part 1, *Econometric Theory* 4, 468-497.

[91] Park, J. Y. and P. C. B. Phillips (1989), Statistical inference in regressions with integrated processes: Part 2, *Econometric Theory* 5, 95-131.

[92] Pere, P. (2000), Adjusted estimates and Wald statistics for the AR(1) model with constant, *Journal of Econometrics* 98, 335-363.

[93] Perron, P. and Z. Qu (2007), A simple modification to improve the finite sample properties of Ng and Perron's unit root tests, *Economics Letters* 94, 12–19.

[94] Pesaran, H. M., and A. Timmermann (1995), Predictability of stock returns: robustness and economic significance, *Journal of Finance* 50,1201–1228.

[95] Petkova, R. and L. Zhang (2005), Is value riskier than growth?, *Journal of Financial Economics* 78, 187-202.

[96] Phillips, P. C. B. (1986), Understanding spurious regressions in econometrics, *Journal of Econometrics* 33, 311–340.

[97] Phillips, P. C. B. (1987a), Time series regression with a unit root, *Econometrica* 55, 277–302.

[98] Phillips, P. C. B. (1987b), Asymptotic expansions in nonstationary vector autoregressions, *Econometric Theory* 3, 45-68.

[99] Phillips, P. C. B. and B. E. Hansen (1990), Statistical inference in instrumental variables regression with I(1) processes, *Review of Economic Studies* 57, 99–125.

[100] Phillips, P. C. B. and T. Magdalinos (2007), Limit theory for Moderate deviations from a unit root, *Journal of Econometrics* 136, 115-130.

[101] Phillips, P. C. B. and T. Magdalinos (2009), Econometric inference in the vicinity of unity, Working Paper, Singapore Management university.

[102] Phillips, P. C. B. and P. Perron (1988), Testing for a unit root in time series regression, *Biometrika* 75, 335-346.

[103] Plosser, C. I. and Schwert, G. W. (1978), Money, income, and sunspots: Measuring economic relationships and the effects of differencing, *Journal of Monetary Economics* 4, 637-660.

[104] Polk, C., S. Thompson, and T. Vuolteenaho (2006), Cross-sectional forecasts of the equity premium, *Journal of Financial Economics* 81, 101–41.

[105] Pontiff, J., and L. D. Schall (1998), Book-to-market ratios as predictors of market returns, *Journal of Financial Economics* 49,141–60.

[106] Roy, A. and W. A. Fuller (2001), Estimation for autoregressive time series with a root near one, *Journal of Business and Economic Statistics* 19, 482-493.

[107] Rozeff, M. S. (1984), Dividend yields are equity risk premiums. *Journal of Portfolio Management* 11, 68–75.

[108] Said, S. E. and D. A. Dickey (1984), Testing for unit roots in autoregressive-moving average models of unknown order, *Biometrica* 71, 599-607.

[109] Saikkonen, P. (1991), Asymptotically efficient estimation of cointegration regressions, *Econometric Theory* 7, 1-21.

[110] Schwarz, G. (1978), Estimating the Dimension of a Model, *Annals of Statistics* 6, 461-464.

[111] Schwert, G. W. (1989), Tests for unit roots: A Monte Carlo investigation, *Journal of Business & Economic Statistics* 7, 147-159.

[112] Seo, M. H. (2006), Improving unit root testing with a new long run variance estimator, working paper, London School of Economics.

[113] Stambaugh, R.F. (1999), Predictive regressions, *Journal of Financial Economics* 54, 375-421.

[114] Stock, J.H. and M.W. Watson (1993), A simple estimator of cointegrating vectors in higher order integrated systems, *Econometrica* 61, 783-820.

[115] Tanizaki, H. (2000), Bias correction of OLSE in the regression model with lagged dependent variables, *Computational Statistics and Data Analysis* 34, 495-511.

[116] Torous, W., Valkanov, R., and S. Yan (2004), On predicting stock returns with nearly integrated explanatory variables, *Journal of Business* 77, 937–966.

[117] Tsay, R. S. and G. C. Tiao (1990), Asymptotic properties of multivariate nonstationary processes with applications to autoregressions, *Annals of Statistics* 18, 220-250.

[118] Tsui, A. K. and M. M. Ali (1994), Exact distributions, density functions and moments of the least-squares estimator in a first-order autoregressive model, *Computational Statistics and Data Analysis* 17, 433-454.

[119] Viceira, L.M. (1997), Testing for structural change in the predictability of asset returns, Unpublished manuscript, Harvard University, Cambridge, MA.

[120] Vinod, H. D. and L. R. Shenton (1996), Exact moments for autoregressive and random walk models for a zero or stationary initial value, *Econometric Theory* 12, 481-499.

[121] Zaman, A. (1996), *Statistical Foundations for Econometric Techniques,* Academic Press, London.